Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

EUROPEAN DOCTORATE IN LAW AND ECONOMICS

Ciclo 27°

Settore Concorsuale di afferenza: 13/A3

Settore Scientifico disciplinare: SECS-P/03

TITOLO TESI

SOCIAL WELFARE AND BEHAVIORAL PUBLIC POLICIES

Presentata da: Marco FABBRI

Coordinatore Dottorato

PROF. LUIGI ALBERTO FRANZONI

Relatore

PROF.MR.DR. FRANCESCO PARISI

PROF.MR.DR. LOUIS T. VISSCHER

Esame finale anno 2015

Social Welfare and Behavioral Public Policies

Marco Fabbri*

Department of Economics, University of Bologna
Hamburg Institute of Law and Economics, Hamburg University
Rotterdam Institute of Law and Economics, Erasmus University Rotterdam

*PhD Candidate and Erasmus Mundus Scholar, European Doctorate in Law and Economics, University of Bologna, Department of Economics, Erasmus University Rotterdam, Rotterdam Institute of Law and Economics, Hamburg University, Hamburg Institute of Law and Economics. *Current address:* Department of Economics, University of Bologna, Piazza Scaravilli 2 40126 Bologna, Italy. Tel.: +39-051-208-8550. E-mail: marco.fabbri@edle-phd.eu. I am deeply indebited with my advisors Francesco Parisi and Louis Visscher, who led and supported me throughout the writing of this work. I am grateful to Emanuela Carbonara for the encouragement and support provided in writing chapters 3 and 4: those chapters would have not been written without her help. I also thank Diogo Gerhard, with whom I coauthored the paper "When Choosing the Social Welfare Function Really Matters: a Quantitative Analysis", *Rotterdam Institute of Law and Economics (RILE) Working Paper Series* No. 2013/01 that constitutes the core of chapter 2, and Sigrid Hemels, with whom I coauthored the paper "Do You Want a Receipt? Combatting VAT and RST Evasion with Lottery Tickets", *Intertax: international tax review*, 2013, 41(8), pp. 430–443, that discusses the legal implications of the ideas presented in chapter 3. This work benefited from insightful suggestions and valuable comments from Paolo Nicola Barbieri, Maria Bigoni, Marco Casari, Robert Cooter, Michael Faure, Luigi Alberto Franzoni, David Gamage, Andrea Geraci, Riccardo Ghidoni, Jonathan Klick, Matthew Rabin, Roberto Weber and conference and workshop participants at the 2014 European Association of Law and Economics Annual Meeting, the 2013 Italian Association of Law and Economics conference, 2013 European Association of Law and Economics Annual Meeting, the 2013 German Association of Law and Economics Annual Meeting, the IX Young Economists' Conference on Social Economics at University of Bologna, the VI IMPRS Uncertainty Topics Workshop at Erasmus University Rotterdam and seminars participants at University of Bologna, Hamburg Institute of Law and Economics, Rotterdam Institute of Law and Economics and University Paris II for helpful comments. I am grateful to the Alfred P. Sloan foundation that provided financial support for the experiment that constitutes the core of chapter 4 and to Stefano Rizzo that provided valuable research assistance. The usual disclaimer applies.

Contents

1. Introduction

1.1. Book Topic, Methodology and Results

The topic of this book is related to the rising field of behavioral public policymaking. Scholars belonging to this movement propose policy interventions that address systematic and predictable violations of rationality to steer agents' behavior in directions that are self-beneficial, possibly without limiting individual autonomy or restricting freedom of choice. The intuition at the basis of behavioral public policies is that humans are characterized by limited cognitive abilities and their choices are often influenced by details not included in models of standard decision-making. Therefore, the policy analyst that is able to identify, explain and predict nonstandard behavioral regularities could make use of this knowledge to promote welfare-improving policies. The interest surrounding behavioral public policies comes from the fact that its proposals are easy to implement, relatively cheap and in many cases respectful of individual freedom of choice.

This book proposes a detailed introduction to behavioral public policymaking and three original contributions. The introductory chapter focuses on specific issues of welfare analysis. Welfare analysis is a two steps procedure. First, the analyst determines how a policy affects individuals' well-being. Second, the analyst aggregates the well-being across individuals. In the remaining sections of the introduction, I focus on the first step of the procedure. I propose an overview of the literature and of the still open debate regarding this issue between behavioral science scholars. I focus on the second step of the welfare analysis procedure in chapter 2. Beside introducing and discussing the problem of aggregation of individuals' utility and proposing an overview of the literature, in this chapter I also suggest a methodological contribution concerning the social analyst's choice of the social welfare function. In chapter 3 and 4 I discuss two innovative policies that make use of behavioral regularities in order to increase social welfare.

While I introduce the reader to the philosophy behind behavioral public

policymaking and I provide a summary of actual applications, possible developments and critiques, the main focus of this book is not to provide a full discussion regarding its merits and the flaws. Instead, my main objective is to contribute to the discussion *within* the behavioral public policymaking movement, suggesting new ideas and original contributions. To investigate the social policy issues object of this book, I employ state of the art methodologies and techniques of economic analysis. Specifically, in chapter 2 I provide a quantitative analysis of the choice of the social welfare function when performing economic analysis of social policy. The problem of choosing a specific form of social welfare function is a key aspect not exclusively of behavioral public policymaking, but of public choice, social choice and welfare economics as well. My goal in chapter 2 is to suggest a methodological advance for this problem by deriving quantitative relationships between a set of social welfare function specifications that aggregate individuals' well-being. To achieve this goal, I formally prove the results and the propositions contained in this chapter using mathematical analysis. I show that, in general, results obtained representing social welfare through a particular combination of individuals' well-being and a method for the aggregation of individual utilities can only be generalized to a subset of the possible social welfare function specifications. Moreover, I highlight under which conditions different combinations of individual well-being representation and aggregation method rank in the same order alternative states of the world. I then derive quantitative conditions under which the policy analysis results could be extended to different social welfare functions. Imposing some restrictive conditions on the redistributive transfers considered, I also demonstrate that it is possible to generalize a set of analysis results. Finally, I show how quantitative conditions necessary to generalize the results obtained assuming particular social welfare functional forms vary when the interest groups affected by the policy have different sizes.

In chapter 3, I discuss the application of a behavioral policy to contrast

indirect tax evasion. Governments both in developed and developing countries are facing the problem of value added tax (VAT) and retail sales tax (RST) evasion. This explains a growing interest in policies alternative to the traditional methods of deterrence. This chapter describes the achievements resulting from a zero cost policy against VAT and RST evasion based on rewards. Customers are encouraged to request an invoice by changing the invoice into a lottery ticket, thereby making VAT and RST fraud and evasion more difficult for suppliers. Such a policy has, for example, been introduced in some Asian countries. My goal in this chapter is to explain the puzzling empirical evidence of the policy success and to propose a model that allows policymakers to predict the outcomes of the policy when applied in specific contexts. The methodology that I employ is a combination of mathematical analysis and of empirical work based on a calibration exercise. After having characterized VAT and RST evasion as a special kind of public good situation, a theoretical model based on non-expected utility theory is presented. Given this theoretical framework, I provide examples based on calibration exercises showing the possibility to predict the policy outcome in different socio-economic contexts. Finally I discuss the possible countervailing effects as well as the positive long-term effects generated by the introduction of the policy.

In chapter 4, I study the effects of social influence on third-parties' decision to engage in costly punishment. My chapter is the first contribution investigating the topic. My goals in this chapter are to isolate and estimate the causal effect of social influence on third-party punishment and to identify the channels through which social influence operates. To achieve these objectives, I first propose a mathematical model of decision-making that includes social influence effects. I test my model predictions setting up a laboratory experiment based on the methodology of experimental economics. I then analyze the resulting data employing state of the art econometric tecniques. The design of the experiment is based on a dictator game. I exclude payoff comple-

mentarity among punishers and I elicit punishment decisions both in isolation and after having provided information regarding actual peers' punishment. I find evidence that the amount of punishment chosen by third-parties is influenced by beliefs about the amount of peers' punishment. Moreover, the larger the difference between third-parties beliefs about the level of peers' punishment and actual peers' punishment, the more likely the third-parties modify the initial punishment decision. I also find that more self-regarding third-parties are less affected by social influence. I then disentangle the effect of normative social influence from that of informational social influence and I show that some subjects are responsive to the former type of social influence but not to the latter. Finally, I discuss the possibility to enact policies that exploit this behavioral regularities.

Before proceeding with my original contributions, in the remainder of this chapter I introduce the reader to the central topic of this book providing a detailed discussion of behavioral public policymaking. Specifically, in the next section, I provide an overview of the concepts underlying behavioral policymaking, I highlight the analogies and the differences with classic public policy analysis and I discuss a set of nonstandard behavioral regularities in individual decision-making that could be exploited in order to enact behavioral policies. In section 1.3 I discuss a fundamental methodological problem of behavioral public policy, that is the choice of a suitable welfare criterion for conducting social policy analysis, and I report an overview of different possible criteria proposed by scholars. In section 1.4 I then summarize the debate between Paternalism and Libertarianism connected to the implementation of behavioral policies and I discuss a third-way that could in principle reconcile these positions, the so called "Libertarian Paternalism" approach. This chapter is concluded by section 1.5 where I discuss the challenges faced by scholars operating in behavioral sciences that are interested in increasing their direct influence on policymakers.

1.2. Behavioral Public Policy[1]

1.2.1. Definition and Main Challenges

"If you see a man with a razor in his hand yelling that he wants to cut his own finger", says a common joke among economists, " then you should help him sharping the blade". What makes the joke (at least for someone) funny is that its counter-intuitive suggestion is derived from a straightforward application of the utility maximization principle based on revealed preferences (Samuelson, 1938), the benchmark commonly used in welfare economics analysis[2]. In fact, welfare analysis in neoclassical economics typically assumes that people reveal preferences through their chosen actions. Therefore, according to this theory, the choices that an individual makes are also those that maximize her utility. Accordingly, the objective of the social planner is to promote interventions aimed at maximizing people's utility, that is satisfying agents' preferences revealed by their choices. Therefore, if for whatever reason the person in our joke prefers having his finger cut, and his action is not affecting anyone else' utility, why should the social planner stop him?

By relying on the revealed preference approach neoclassical welfare economic analysis does not distinguish between individuals' choices and well-being. To be more precise, the neoclassical policy analyst infers the nature of well-being from the action chosen by individuals and he acts like an individuals' proxy, deriving their policy choice from observed actual consumption choices in similar situations.

Neoclassical economic models typically assume that people make decisions following the principle of rationality[3]. Broadly speaking, a rational agent makes decisions as if he would be able to consider and process all the available information, to engage in cost-benefit evaluations and to smooth present and

[1] This section is mostly based on materials discussed in Bernheim and Rangel (2012).

[2] I will discuss in details both the utility maximization and the revealed preferences approach in section 1.3.

[3] I discuss the key assumptions of rationality in subsection 1.3.2.

future consumption according to his expectations. As a consequence, given preferences, constraints and available information, the agent would end up making the choice that guarantees him the maximum (expected, if we talk about future outcomes) well-being[4].

However, in the last decades, experimental psychologists and behavioral economists documented that in some situations of great economic relevance, individuals systematically depart from economists' neoclassical assumption of rationality (for an overview see for example Della Vigna, 2009, Kahneman, 2003 and Akerlof and Shiller, 2010). Researchers found that decision-making processes are the result of two coexisting and interacting mental systems: an impulsive, short-term focused one ("System 1") and a reflexive, long-term oriented one ("System 2") (see Kahneman, 2011 for a discussion of this point; see Hsu et al., 2005 for a contribution that identifies the neural correlates responsible of the activation of different areas of the brain connected with the two systems). While decisions made by System 2 would be fairly consistent with neoclassical economic predictions, nevertheless the influence of System 1 is responsible for the aforementioned biases (Loewenstein and Haisley, 2007). The problem with System 1 is that, when people make decisions on the basis of emotions, neglecting information or attaching exagerate weight to the present, they might end up making choices contrary to their own self-interest. For example, they could take excessive risks, make decisions that they will later regret or forego possibilities of high future gains in order to avoid small immediate costs (Camerer et al., 2004).

For the purposes of the present work it is important noticing that the direction of the aforementioned deviations from the optimal behavior are often predictable. As a consequence, scholars have been able to produce models of decision-making that incorporate these regularities (Ariely and Jones, 2008; Camerer and Loewenstein, 2004a). Exploiting these predictions, behavioral

[4]Economists usually employ the concept of "utility". I will discuss this concept below.

policymakers have suggested policies to counteract biases and redirect patterns of behavior that usually hurt people to enhance the optimality of decision making (Loewenstein and Haisley, 2007). Therefore, behavioral policies in principle aim at correcting patterns of behavior that produce "suboptimal" outputs and redirecting them to alternative choices that make people "better off". However, what is the "optimal" choice and what does it mean to make someone "better off"? It is obvious that behavioral policies cannot rely on the preference-based criterion for optimality. In fact, as we have mentioned utility maximization theory relies on the assumption that whatever action an individual voluntarily chooses must be welfare-enhancing. Thus, it does not make sense to evaluate if an agent is making a suboptimal decision using a benchmark measure built on the premise that people always make optimal choices.

However, recognizing that individuals might not choose what they want creates problems with respect to the identification of a suitable welfare criterion. So far, among behavioral economists no consensus regarding standards and criteria to adopt has emerged. Broadly speaking, it is possible to identify two schools of thought. On the one hand, in the opinion of some scholars policy evaluations must maintain a strict adherence to the doctrine of revealed preferences. According to this view, observed "anomalies" in individuals' decision-making should be explained by an extension of the preferences domain, as for example in Gul and Pesendorfer (2001, 2004).

On the other hand, other researchers investigated the possibility to relax, modify or depart from the principle of revealed preferences in conducing welfare analysis. A number of proposals have been advanced in in this direction and in section 1.3.3 I report a detailed overview of major contributions. A common distinctive tract of all these proposals is the division between a *positive* analysis of a policy effects and a *normative* evaluation of the well-being. This division allows behavioral policy analysts to engage on issues of great social importance. They can, for example, meaningfully address the

questions raised by self-destructive behaviors (considering the example at the beginning of this section, would you take seriously the policy prescriptions of someone that suggests sharping the blade?) or make a sense of the claim that the average household saves "too little" for retirement.

However, departing from the revealed preference approach creates also serious concerns. In principle, this approach guarantees individuals' freedom of choice and it protects their choices against a-priori condemnations. Nonetheless, once this approach is abandoned, governments become entitled of the possibility to condemn individuals' choices and to set "beneficial" restriction of personal freedom. Therefore, given this danger, the determination of precise standards of evidence for departing from the principle of revealed preferences and the determination of a normative welfare criterion acquires a fundamental importance in behavioral public policymaking. I devote section 1.3 to an investigation of this problem and a review of the literature this area. Moreover, I provide in section 1.4 an overview of the debate between supporters of paternalistic interventions aimed at counteracting behavioral biases and defendants of liberalistic policies based on individual freedom of choice. I conclude this introduction discussing in section 1.5 some of the practical challenges that scholars in behavioral sciences face when they have to transplant results of their research into the political arena and the policymaking process.

Before proceeding to the next section, below I revise some common behavioral regularities that can be counteracted and used by policymakers. I will delay the discussion of overweighting of small probabilities and social influence respectively to chapter 3 and 4, where I suggest two behavioral policies based on these nonstandard behavioral regularities.

1.2.2. Occasions for Behavioral Public Policy Interventions

Immediate Feedback, Reinforcement and Default Rules

It is perfectly natural (and consistent with standard economic model predictions) that people evaluate the same good more if consumed today than

tomorrow. In economists' words, people discount the utility obtained from future consumption. Hence, for example an agent facing the choice to receive x money right away or to wait and receive (x + y) money a month from now will accept the later option only if y is positive and large enough. However, how large must y be for the agent to choose the waiting option? In principle, there is not a correct answer: people have heterogeneous preferences ("time preferences" in this context) that could vary a lot across individuals. An impatient person for choosing to wait will require a high y compared to a more long-term oriented person. Similarly, for a given y, different individuals might maximize individual utility either consuming the good immediately or waiting, according to their time preferences.

However, even assuming that an agent could be extremely impatient, in practice a consistent fraction of people systematically fail to make decisions involving negligible short-term costs and huge long-term gains, showing a behavior that is hard to justify from any reasonable perspective (O'Donoghue and Rabin, 2001). These choices often represent important determinants of the welfare of the decision-makers themselves, as in the case of saving for retirement or investment in cost-saving technologies. Even worse, these choices could be detrimental for people's own health, as in the case of smoking (Volpp et al., 2009), obesity (Jeffery et al., 1983) or drug addiction (Higgins et al., 2000).

What is most strinking about the behavior of people making these decisions that go against their own self-interest is that often they *want* to select the opposite option. In situations like the one just described, people show to be affected by self-control problem and "present-biased preferences" (Benhabib et al., 2010). These agents "hyperbolically discount" future consumption, trading off the possibility of high future welfare gains in favor of an immediate but small benefit (Laibson, 1997). Moreover, if asked about their future plans regarding the same situation, people report that they *will* modify their actual behavior (and, of course, when the future becomes the present, they

fail to do it). These time-inconsistent behavior is generated by the so called "projection bias": people fail to recognize that their future selves will have the same present-biased preferences when the time to make a decision comes. The result is an endless procrastination and failure to modify the status quo, even when inertia generates clearly suboptimal outcomes (O'Donoghue and Rabin, 1999).

In all the situations described above, behavioral policy interventions might help individuals affected by behavioral biases that would like but are unable to make welfare-improving decisions, without reducing other people's freedom of choice. For example, researchers document that small but frequent financial incentives conditioned to compliance with some predefined behavior are significantly effective in reducing drug addiction (Higgins et al., 2000) and investments in the use of fertilizer, that in turn greatley increase subsequent harvest (Duflo et al., 2011). Moreover simply reminding on a regular basis people about the opportunity to save for retirement increases the saving rate (Karlan et al., 2010).

Finally, in light of the effects of inertia and procrastination, default rules acquire a special importance (Johnson and Goldstein, 2013). Our lives are complex and we have to make hundreds of decisions everyday, most of which - even if important - we do not know enough about or we do not pay attention to. As a result, people often tend to avoid choosing and stick with whatever default option has been selected (if you are reading these lines on a computer screen or have your cellphone in your pocket, think about how many of the hundreds of default options that were already set for you when first bought these items have been modified!). This "yeah, whatever!" behavior is quite common for people in many decisions, and setting a default rule that helps achieving welfare-improving behaviors could benefit naive or careless individuals, allowing the others to freely select their preferrred option (Dinner et al., 2011).

Furthermore, in many situations deciding what is going to happen if people

fail to make an explicit choice is simply not avoidable. For example, it is well known that in the US a substantial fraction of workers fail to choose the rate of saving for retirement for the current year (Thaler and Sunstein, 2008, pp.40-52). What should the employer or the state assume in this case as a default? One option is to confirm the previous year's saving rate. A second possibility would be to assume that the worker saves nothing. From the above discussion about the power of inertia and procrastination, it is evident that the two default options would have very different consequences for the aggregate level of pension savings and future welfare, even if in principle workers could freely and in any moment opt-out from the default plan.

Loss Aversion

People hate losses more than they love gains. Roughly speaking, gaining something increases a person's utility by half of the amount losing the same thing makes the person worse-off. Researchers call this behavioral regularity "loss aversion". Attempting to explain this empirical evidence that clashes with the prediction of standard economic rationality, Kahneman and Tversky proposed the famous theory of decision under risk and uncertainty that they called "Prospect Theory" (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992). According to Prospect Theory, in decisions involving probabilistic outcomes agents assume their status quo as a reference point and they attach higher weights to changes happening in the domain of the losses than to changes involving gains.

Therefore, in many choices involving risk and uncertainty, loss aversion could exacerbate the problem of inertia and procrastination: agents are reluctant to modify the status quo even when the choice involves gains with high probability and a loss otherwise because of the fear to worsen their actual position (Thaler and Sunstein, 2008). However, behavioral policies could counteract the effects of loss aversion and even exploit them for helping agents increasing their welfare. An example of such a policy is reported in Volpp et al. (2008). The authors run a field experiment with people that voluntarily

enroll in a program for losing weight. In this program, participants have to deposit a certain amount of money in an account and the experimenters match this amount. However, participants must sign a contract stating that they could withdrawn the money only if their weight, measured weekly, falls below a predetermined threshold that implies losing one pound per week. Consistently with non-experimental data coming from similar experiences in different frameworks (Mann, 1972), treated participants lose significantly more weight than participants assigned to a control group where they simply pay for losing weight[5].

Framing Effects

Researchers repeatedly show that framing a decision situation, modifying some apparently uninfluential details, has a tangible impact on people's choices (Tversky and Kahneman, 1981). So for example, labelling a Prisoner's Dilemma game[6] "The Wall Street game" or "The Community game" significantly modifies the level of cooperation (Liberman et al., 2004; Rege and Telle, 2004). So, why not framing situations in a way that redirects agents toward desirable outcomes?

Bertrand et al. (2006) report evidence that framing effects are a powerful tool to increase the take-up into an health care spending account sponsored by employers that is beneficial for employees. In a field experiment sponsored by

[5]There is plenty of funny anecdotical evidence about the power of loss aversion. For example, the captain of the Italian national volleyball team in the ninties, world-champion for twelve consecutive years and considered the best national selection of all times, when asked about the secret for winning so much replied: "Well, it's very simple, the other teams play to win, instead we play to avoid losing." (Notice that in volleyball even results in a match are not possible...).

[6]The Prisoner Dilemma is a well-known and widely used game in social sciences. Two players have to make a simultaneous decision among two possible choices, cooperating or not-cooperating, and each player's payoff depends on his individual choice and on the choice of the other player. While from a social welfare perspective the optimal outcome would be the cooperation of both players, however each player has incentives to free-ride on the cooperation of the other player. Therefore, according to standard game theory predictions, the unique Nash equilibrium of the game is that players do not cooperate.

a large telecommunications company, women employees were invited to view a 15-minute videotape providing information about the importance of taking a mammography. They were randomly divided in two samples: half viewed a video called "The Benefit of getting a Mammography" while half where exposed to a video titled "The Risks of Neglecting a Mammography". While the information contained in the videotape was essentialy the same, framing the decision to take a mammography as an opportunity to avoid a potential loss induced a significantly higher percentage of employees to do it in the subsequent month with respect to the sample of employee exposed to framing as a gain. Studies on framing effects obtained similar results in situations concerning people's decision to subscribe insurance policies (Johnson et al., 1993) and to contribute to charities (Davis et al., 2005).

Goal Gradients

Runners and bike racers sprint with renewed energies when approaching the finish lane. Default rate on mortgages drops almost to zero when the final total repayment is close. Students drop out rate falls when the last exams are approaching. And PhD candidates become able to work unbelievable number of hours when the dissertation deadline is approaching (procrastination plays a role too here). In fact, proximity to the final objective increases motivation (Kivetz et al., 2006) and people often fail to achieve a goal because they feel "stuck in the middle" (Bonezzi et al., 2011; Hull, 1932).

Given this evidence, why do not reducing the distance between the starting point and the finish lane, splitting the total distance in multiple shorter starts-and-arrivals[7]? Many microfinance institutions use this principle to reduce default rate on loans (Morduch, 1999). The same principle is also at the basis of Individual Development Accounts (IDAs). IDAs are savings ac-

[7]This tecniques is applied on a regular basis among extreme hikers, including the person that is writing: if one focuses on going through the 20 miles and 13000 feet of altitude gap of the Pico Tarquino in Cuba within one day, he would never even consider to get out of bed.

counts in which the state or a selected organization matches the deposits with the objective to help low income families to either buy their own house, open a small business or invest in children education (Schreiner and Sherraden, 2007). IDAs have proved to be successfull in increasing poor people savings not only in developing countries: in a field experiment conducted in Oklahoma, Mills et al. (2008) found that in a 4-years time horizon participants assigned to the treatment where they could open an IDA account significantly increased both savings amount and the probability to become houseowner.

Moreover Loibl et al. (2012) report preliminary results of a series of experiments where the authors test different strategies to increase savings and retention rates in IDA programs. The preliminary evidence suggests that, holding constant the total cost of matching, in the IDA programs where the match rate increases overtime saving deposits are higher than those where the match remains fixed.

1.3. A Fundamental Methodological Problem: What Welfare Criterion?

In the section above I discussed behavioral policies based on the presumption, supported by empirical evidence, that individuals' actual choices are not always welfare-improving. It has been underlined that a fundamental methodological problem for behavioral policymaking is determining which welfare criterion to embrace. Scholars have proposed different concepts of welfare suitable for behavioral policy analysis. I discuss in details some of these criteria in paragraph 1.3.3. However, before presenting welfare criteria that depart from the revealed preferences approach, in the next paragraph I report an historical overview of the schools of thought in welfare economics and of the non preference-based concepts of utility proposed.

1.3.1. Schools of Thought in Welfare Economics

A fundamental step in the economic analysis of social policy is concerned with evaluating the desirability of policies effects, that is, to produce norma-

tive statements. This normative analysis is the core of welfare economics, the branch of economic science evaluating well-being from the allocation of productive factors in terms of economic efficiency and desirability or resources allocation. Scholars over the years embraced different positions and enriched the debate concerning a series of methodological and normative issues regarding welfare analysis. In the following paragraphs I provide an overview of the different schools of thought.

Welfarism

According to social welfare philosophy (or utilitarian philosophy), it is necessary to evaluate different states of the world in terms of their end-state distributional results. In fact, the purpose of welfare economics is to obtain a social ordering over alternative possible states of the world, thus promoting, when possible, welfare-improving policies. In order to achieve this social ordering, welfare economics embraces well defined normative principles[8]:

1. The utility principle: rational individuals maximize their welfare by ordering and choosing the preferred option
2. Individualism: individuals are the only judges of what contributes most to their utility
3. Consequentialism: utility is derived only from the outcomes of behavior and processes
4. Welfarism: the goodness of any state of the world could be judged only by the level of utility attained by individuals in that situation

The theoretical basis for the determination of individuals' ordering over alternative states of the world is the utility maximizing theory of consumer choice, whereby consumers rank alternative states according to a set of preference or-

[8]In the following paragraph, I only provide an overview of the normative foundations of welfare economics. For an in-depth discussion see Arrow, 1951; Harsanyi, 1977; Sen, 1997

derings. The concept of utility[9] has been for a long time the center of debate within welfare economics scholarship. Two possible main interpretations are commonly assumed: *preference satisfaction* and *hedonic welfare* (Van Praag, 1993). Hence, utility includes any element affecting individual preference satisfaction or welfare, expressively excluding any other element. The process of utility measurement assigns numerical values to different bundles, goods or states of the world, representing an individual preference ordering among alternatives.

The individualism principle restricted the source of utility in welfare economics to individual judgements and to goods and services that he himself consumes[10]. This principle implies for the policy analysts operating in the framework of welfare economics to discard any other aspect not affecting individuals' utility.

The consequentialism principle states that the focus of the analysis must be restricted to outcomes, excluding from the analysis any other element. Some attempts have been made to enlarge this principle in order to include additional factors other than outcomes, notably processes and procedures (see Birch et al., 2003 for an example and further references). Nonetheless, this further enlargement only considers processes and procedures affecting individuals' preferences and so individual utility.

Finally, the welfarism principle directly links social welfare to the utility of individuals, excluding from social analysis considerations not directly linked

[9]If there is uncertainty about the future, economists talk about "expected utility". In this case, the representation of preferences refers to a more elaborate theory compared to the case of certainty about the future, the Expected Utility Theory proposed by Von Neumann and Morgenstern (1944). I will discuss in chapter 3 the Expected Utility Theory. In fact, in some situations social decision results could be affected by performing the analysis before the resolution of uncertainty or after (Myerson, 1981). However, in order to avoid technical complications, since the focus of my work is on a different topic I do not consider the possibility of uncertainty about the future in the discussion presented in this chapter.

[10]It is however common among welfarist economist to enlarge the source of utility also to some other dimensions. See for example the analysis of Becker (1968a); Becker and Becker (2009) on the family and marriage or Culyer (1971) on health care and donations.

to individual utility.

Welfare economics scholars commonly assume also anonimity. According to the anonymity principle, each individual's well-being affects social welfare in a symmetric manner. This implies that under the framework of welfare economics, no individual or interest group's achievements of well-being can be considered more important than those of others. As in the cases of individualism and consequentialism, the anonimity principle has been sometimes superated. Indeed, some scholars argue that utility functions weighting differently subjects endowed with unequal level of utility reflect individuals' preferences for redistribution (see for example Roemer, 1998).

The traditional welfare economics stream is based on the above tenets. Following Sen (1977), we will call this school "welfarist economics", and we will point out its discrepancies with the recent development of extra-welfarism in the next paragraph. Within the framework of welfarist economics, some divisions mirror different approaches to some specific point. The "classical" welfarist economic school considered utility cardinally measurable and possible to be added and compared across individuals (Diamond, 1967). Hence, maximizing the sum of individual utilities constitutes optimality and the objective of the social analyst is to achieve "the greatest happiness of the greatest number that is the measure of right and wrong" (Bentham, 1776, Preface, p.ii). Economists nowadays tend to considered the idea of cardinal utility sorpassed. However, in specific contexts like decision making under risk and welfare analysis cardinal utility comparison are sometimes still employed (Köbberling, 2006).

Conversely, the "neo-classical" school questions the possibility to measure utility cardinally. As a consequence, scholars of the neoclassical school in their welfare analysis consider an ordinal measure of utility.

Neoclassical welfare analyisis is based on individuals' preferences, that are revealed by means of chosen actions[11]. Once individuals' preferences are known, it is subsequently possible to aggregate them using the Pareto principle. The Pareto principle implies that a social state X has to be preferred to an alternative Y if every agent is at least as well in X as in Y, and at least one agent is better off in X. Up to this point, the policy analist could produce welfare evaluation without having to engage in any "unscientific value judgement"[12] (Fleurbaey and Hammond, 2004).

However, this procedure presents two problems. First, it is not per se possible to identify one desired state of the world from the set of all possible Pareto states. A second problem comes from the fact that most of the policies that modify the status quo make some agents better-off while reducing the utility of others. The solution to these problems adopted by many welfare economists is to weight and then compare the utilities of different households according to some ex-ante value judgment. However, it should be noted that some distinguished scholars question the possibility to produce meaningful interpersonal comparison of individual utility (See for example Robbins, [1932 Or. Ed.]-2007; Buchanan, 1959; Buchanan et al., 1979 and Boadway and Bruce, 1984). I do not enter here in the highly debated question regarding the possibility of interpersonal comparison of utility. I redirect the interested reader to the wide strand of literature in economics and other social sciences, for example Baron (1993); Binmore (1989); Harsanyi (1980); Hammond (1991); Luce (2010); Sen (1973).

Assuming that interpersonal comparison of utility is possible, two criteria have been developed for verifying whether, under the assumption that the gainer is able to compensate the loser, a modification of the status quo leads the economy toward Pareto-optimality: the "Kaldor criterion" and the

[11]The next paragraph is devoted to a in-depth analysis of this point.

[12]To be more precise, the only value judgement implied by this analysis is that people's preferences could be estimated from the observation of the choices they perform.

"Hicks criterion". The Kaldor criterion states that a change in the status quo would increase Pareto efficiency if the gainer is ready to pay an amount greater than the minimum amount the loser is willing to accept in order to achieve this change. By contrast, the Hicks criterion states that if the loser is prepared to offer a maximum amount smaller than the minimum amount the loser is willing to accept in order to prevent the status quo modification, the change increases Pareto-efficiency. At this point, the social planner intervention is justified through the concept of asset egalitarianism, introduced by Arrow (1973). The asset egalitarianism principle considers society's assets as part of a unitarian common wealth of humanity, the social welfarists' object of maximization. Therefore, a change in the status quo through redistributive policy becomes desirable if it raises social welfare.

Now, I have to stress here that welfare results derived from this analysis depend on the aforementioned value judgements. In fact, these value judgments are implicitly reflected in the specification of the analytical tool employed by social analysts in order to perform social policy analysis, namely the social welfare function. I devote chapter 2 of this dissertation to discuss the importance of social welfare functions for policy analysis, so I redirect the reader there for definitions and discussions of concepts.

Before turning to the next paragraph, below I report an overview of the extra-welfarist schools in welfare economics. Extra-welfarism is expecially important for the purposes of this work, since the welfarist tradition bases its welfare analysis on the theory of revealed preferences, a position that contradicts the fundamental idea of behavioral policymaking.

Extra-welfarism

The tenets on which welfarist economics is founded have been considered too restrictive by some scholars working on the economic analysis of social policy. Early articles introduced concepts like merit goods (Musgrave, 1959), specific egalitarianism (Tobin, 1970), spheres of justice (Walzer, 1983), basic goods (Rawls, 1971) and capabilities (Sen, 1980; Sen et al., 1993), whose the-

oretical foundations lie outside the welfarist approach. In recent years, this school of thought was named "extra-welfarism" to emphasize the distinction with traditional welfarist economics. Extra-welfarism has been gaining particular importance in health economics literature (Culyer, 1989). According to Brouwer et al. (2008), welfarist economics is a special subset of extra-welfarism where the evaluative space is narrowed down, and both schools fit within the (widely interpreted) paradigm of welfare economics. In an attempt to draw a distinction between welfarist economics and extra-welfarism, the authors point out four main factors.

First of all, the source of evaluation of utility is not anymore confined to the individual affected by the social choice, but includes also expert third parties and representative samples of the general public not directly affected. Second, extra-welfarism explicitly allows attaching different weights to distinct individuals. Specifically, the policy analyst could paternalistically introduce ethical considerations, most notably related to equity and justice. Third, the interpersonal comparison of the relevant outcomes is explicitly allowed. However, differently from welfarist economics, cost-benefit (or cost effectiveness) analysis does not only compare utilities, but also capabilities and other characteristics according to the field of application. Finally, extra-welfarism includes as a relevant outcome individual utility and other possible measures of well-being, that could be selected by the analyst according to the specific field of interest. Hence, the choice of the analyst becomes explicitly normative and acquires further importance compared to the welfarist framework. This point acquires the highest importance in light of the discussion about the choice of a welfare criterion for behavioral public policies, since a strict application of the preference-based criterion commonly assumed in welfare economics is not a suitable option.

In the next paragraph I focus on the assumptions implicit in neoclassical welfare economics analysis. I will make this assumptions explicit and underline their implications for the policy analyst. In the following section I

present instead an overview of the approaches proposed by the behavioral literature in the attempt to proceed with a welfare analysis that allows for the relaxation of these assumptions.

1.3.2. The Neoclassical Welfare Analysis and its Assumptions

In the previous paragraph, we mentioned how neoclassical welfare analysis is based on the principle of individualism: what is good or bad for society reflects what is good or bad for the individuals belonging to the society. Therefore, this principle guides the policy analyst in the comparison and choice among alternative policy options. The analyst is indeed supposed to suspend his individual value judgment and act as each individual's proxy (Bernheim and Rangel, 2005, pp. 5). In this paragraph we focus on the meaning of "act as each individual's proxy".

For a given set of conditions, the neoclassical policy analysts derives which policy choice to make from the observation of private consuming choices made by individuals. Indeed, standard consumer theory allows to extrapolate public policy outcomes from the observation of private choices. Therefore, the analyst discovers the preferences of the private individual through the observation of her chosen actions. A common way of interpreting the neoclassical approach is that people have well-defined preference rankings and these rankings form the basis for welfare analysis. Following Bernheim and Rangel (2012), we can say that this approach is based on some key assumptions:

1. *Coherent preferences:* each individual has coherent and well-behaved preferences.
2. *Preference domain:* the set of state-continget consumption paths that an individual exibits during his life constitutes his preferences domain.
3. *Fixed lifetime preferences:* individuals do not change overtime or across states of the world the rank order of lifetime state-contingent consumption paths[13].

[13]It is worth noticing that this assumption does not rule out the possibility that pref-

4. *No mistakes:* Each individual always choose the preferred option among the feasible ones given his choice set.

In the next paragraph I consider these assumptions one by one. I report scholars' attempt to relax them and to identify a suitable welfare criterion for behavioral public policymaking.

1.3.3. Relaxing the Neoclassical Assumptions

Relaxing Coherent Preferences

The assumption of coherent preferences implies that people's decision are well-defined and that they are not influenced by irrelevant details or by the context in which they are taken. However, starting with the pioneering work of Tversky and Kahneman (1981, 1986), behavioral scientists show that observed choices are highly context-dependent and that framing greatly influences individuals' decision. Given these observations, some scholars proposed welfare criteria that are not anymore based on the notion of allocation of resources. These contributions introduce a sharply separation between positive model describing choices and normative models describing welfare.

Along this line is developed the capabilities approach is first advocated by Sen (1985, 1999) and developed by (Nussbaum, 2001). This approach rejects the standard preference-based measurement of welfare on the basis of the concept of hedonic adaptation: people adjust individual preferences and expectations to social conditions and to the surrounding environment. Therefore, choices made by agents in a specific situation might not just reveal individual preferences but instead could show that people adapted their preferences to the specific circumstances in an attempt to achieve feasible goals. Therefore, Sen

---

erences ranking changes with age or with some other factors (e.g. mood). However, the assumption states that, holding constant the state of the world, an individual would keep considering the choices made at any specific time as welfare-maximizing. In the same way this assumption implies that an individual in a certain state of the world (e.g. when he is happy) would keep considering as welfare-maximizing the choices made when he was in a different state of the world (e.g. when he was depressed).

and Nussbaum argue in favor of a normative theory of welfare that is based on what people are capable of achieving given surrounding social conditions and the opportunity offered to them. Nussbaum goes further proposing a set of fundamental human capabilities on which this theory should be based.
A notion of welfare based on opportunities and that share some common points with the capabilities approach is the one advocated by Sugden (2004). In his contribution, Sugden formulates a rigorous welfare criterion that justifies the use of opportunities as a welfare standard.
Both the capabilities approach and the opportunity criterion solve the problem of hedonic adaptation and overcome the revealed-preference theory assumption that choices are always welfare enhancing. Nonetheless, these criteria create for the policymaker the problem to determine which capabilities or opportunities must be valued. Despite this critique, I nonetheless discuss in the next chapter how it is often unavoidable for policymakers to engage in some sort of value judgements when performing policy evaluations.

Relaxing Preference Domain

It is possible to identify two classes of behavioral anomalies that are inexplicable through the neoclassical approach but that allow for a welfare analysis if one extends the preference domain. The first anomaly involves temptation and self-control, the second is constituted by social preferences.
Empirical evidence suggest that in a variety of situations individuals engage in time-inconsistent choices and that they rely in various form of precommitment (Ameriks et al., 2007). The solution proposed by Gul and Pesendorfer (2001) consist in defining the preference domain over both allocations and choice sets. If individuals are sophisticated and can correctly forecast the effect of future temptations, they could prefer to constrain future alternatives even when constraints should not have any impact on actual choices. For example, a sophisticated individual wanting to save for the Christmas period could correctly forecast his inability to avoid shopping during the summer sale season. Therefore, she could prefer a commitment device that limits her

future choice set. For example, she could opt for a special saving account that, holding constant the benefits offered, additionally imposes the payment of a penalty for money withdrawn before the month of December[14].
It is important to notice that the Gul-Pesendorfer framework however does not imply a depart from the revealed preference approach. Indeed, individuals maintains, as in the neoclassical framework, the same lifetime preferences ranking at every moment in time (e.g. absent the penalty for withdrawn, the individual would recognize as welfare-maximizing the decision to shop on the summer sale, and she explicitly imposes a constraint because she understands the value of temptation). Therefore, welfare evaluation could be performed by discovering the revealed preferences, assuming that the policy analyst imposes a suitable structure on the choice data.
Behavioral anomalies within the class of social preferences include sharing allocations even absent reputation or reciprocity (see Engel, 2011 for a meta-analysis of the Dictator game), the equality concerns (e.g. Fehr and Schmidt, 1999) and the conformity and social influence effects (for a literature review see Cialdini and Trost, 1998 and chapter 4 of this work). Behavioral economists proposed models where individuals' preferences are defined both over their own and other individuals' consumption bundles. Again, this procedure does not imply abandoning the revealed preferences approach: once a suitable structure is imposed on consumption data, the policy analyst can infer individuals' preferences by observing their consumption choices

Relaxing Fixed Lifetime Preferences
The aforementioned evidence of time-inconsistent behavior and various forms of precommitment motivate also the relaxation of this assumption. Broadly speaking, scholars have adopted two modelling strategies. One possible strategy consists in endowing individuals with well-behaved lifetime preferences

[14]These kind of special bank account in recent years registered an exponential growth in the US.

that vary at different points in time (Laibson, 1997; O'Donoghue and Rabin, 2001). Alternatively, one can allow lifetime preferences to be different across states of nature (Loewenstein, 1996; Loewenstein and O'Donoghue, 2004).
Once these preferences has been measured, then in order to conduct welfare analysis the policy analyst has to aggregate them. Aggregating these preferences within a single individual requires a procedure similar to the aggregation of preferences in a multi-agent situation (indeed, here the modelling strategy implies that we aggregate over "multiple selves"). A branch of the literature exploits this analogy (e.g. Laibson, 1997). Since the introductory section of chapter 2 in this book is devoted to the analysis of the aggregation problem in welfare analysis, I redirect the interested reader there for a detailed discussion of the topic. Another branch of the literature instead proposes to base welfare analysis on the selection of reasonably stable components of preferences (O'Donoghue and Rabin, 1999). Bernheim and Rangel (2012) provide a formal justification for the use of this criterion.

Relaxing No Mistakes

Evidence that preferences and choice diverge motivate a relaxation of the fourth assumption. First of all, there are cases where almost everyone agrees that individuals do not necessarily make choices following their own self-interest, as in the cases of children or agents that are affected by serious mental disorder. More generally, any of the behavioral anomalies we mention as a motivation for the relaxation of the first three assumptions could justify the relaxation of the fourth.
However, assuming that people do not choose what they prefer raises several problematic issues. As we mentioned above, abandoning revealed preferences in favor of some alternative generic normative criterion might entitled governments to interfere with individuals' freedom. Therefore, a first major challenge consists in setting precise criteria and standards for abandoning the revealed preference approach. The literature on this topic is still in its infancy. An interesting proposal is advanced by Bernheim and Rangel (2004).

The authors propose to use findings and advances in applied psychology and neurosciences in order to establish evidence of errors in the brain process mechanisms.

A second issue concerns the identification of preferences. Two basic approaches are possible. Some scholars propose to identify preferences using choice data through an estimation of structural models that incorporate behavioral assumptions on the decision-making processes (see for example Laibson et al., 2007). This process might sounds odd at a first glance: how is it possible to falsify the revealed preferences principle using choice data only? However, this models test the hypothesis of no mistakes jointly with the hypothesis regarding the structure of the decision-making processes that are implicit in the model. Therefore, any evidence of discrepancy between preferences and choices hold as long as the specific non-choice evidence used to motivate the behavioral assumptions of the model hold.

A second approach for the identification of preferences consists in combining choice and non-choice data. One possibility advanced first by Kahneman et al. (1997) is to measure individual well-being on the basis of self-reported evaluations of happiness. Kahneman names this approach "experience utility" as opposed to the "decision utility" usually embraced in economics based on revealed preferences. Experience utility has received significant attention by economists and in recent years there have been important methodological advances regarding the possibility to implement this measure for welfare evaluations (see for example Kahneman et al., 2004; Frey and Stutzer, 2010). In particular, some scholars propose to use this measure of utility in the context of policy evaluations and for identifying appropriate societal trade-off (Layard and Layard, 2011; Bruni, 2007). Moreover, some scholars even argue that happiness should consitute the main goal of policy (Duncan, 2013). From the perspective of behavioral policymaking, happiness measures of welfare have the advantage to be independent from individual choices. Therefore, people's self-reported happiness as a consequence of different choices made

could be employed as a criterion for steering behaviors toward the happiness-maximizing alternative.

However, happiness as a welfare criterion presents several problems (see Loewenstein and Ubel, 2008 for a detailed discussion of each of the following points). First, people seem to adapt relatively quickly to circumstances and to set the reference point for happiness evaluation accordingly. Empirical evidence show that people suffering from permanent disabilities place a high value to their health loss but do not show significant differences in the happiness level if compared with a control sample of non-disable people (Ubel et al., 2005). Hence, measures of welfare grounded on experience utility would suggest policies that fail to capture people's preferences. Moreover, happiness measures are extremely sensible to a wide range of non-normative and volatile factors, such as the happiness of surrounding people, states of mind, emotions or weather conditions (Kahneman and Krueger, 2006). These problems question the possibility to use happiness as the welfare criterion for policy analysis.

In this section I discussed in details some of the technical challenges and problematic issues of behavioral public policymaking. Scholars did not reach a consensus yet on many point and the research agenda is still open. In fact, as I mentioned before, there are significant political danger in abandoning the revealed preferences principle. In the next section I move the debate from the technicalities of the welfare analysis to the broader political debate regarding welfare evaluation.

1.4. Paternalism and Libertarianism[15]

After decades of discussions, the debate between supporters of interventions limiting individuals' action space in potentially self-damaging situations and opposers who instead defend the value of agents' freedom of choice still permeate the political arena. The former view is known as "paternalism", the behavior by a person, organization or state which limits some agent's liberty or autonomy for their own good (Dworkin, 2010). Supporters of paternalism advocate interventions claiming that often agents face important decisions without having sufficient information or the ability to foresee the consequences of their choices (Conly, 2012). For example, smokers might not be fully aware of the long-term consequences of their behavior and workers tend to save too little for retirement possibly because they are unable to correctly foresee their future needs.

Moreover, even assuming full information and perfect capability of prediction, often humans are unable to self-control and self-direct themselves toward what consitutes the "optimal" decision. Hence, a smoker could be perfectly aware of the risks connected to smoking and willing to quit but nonetheless being unable to resist the temptation of the immediate pleasure given by a cigarette (O'Donoghue and Rabin, 2003). Or a worker might want to subscribe to a pension plan that implies high savings but instead he sticks to the default pension plan with minimal savings offered by his firm because of reticence to modify the status quo or simply because he does not pay attention (Chabris and Simons, 2011; Thaler and Benartzi, 2004).

According to the paternalistic view, in these situations a social planner is entitled to promote policies that mandates agents' self-serving choices, or

[15]The still ongoing debate between supporters of paternalism and defendants of libertarianism generates an endless amount of contributions. For the purposes of the present work, in this introduction I will only provide a limited overview of these contributions and I will devote a sizeable part of the discussion to forms of "soft" or "libertarian paternalism" as defined below in this section. For an in-depth literature review of arguments pro- and contra paternalism in its broadest acceptation see the recent article of Sunstein (2013).

the choice people would undertake had they been able to pursue their own self-interest. Paternalism has a long traditions and legal provisions containing paternalistic elements permeate legal systems of modern societies (Burrows, 1995).

Conversely, opponents of paternalism support the principles of "libertarianism". Libertarianism includes "any political position that advocates a radical redistribution of power from the coercive state to voluntary associations of free individuals" (Long, 1998, p.304)[16]. According to libertarianism principles, any paternalistic intervention aimed at increasing agents' own utility by reducing their freedom of choice can only have detrimental results for welfare. In fact, advocates of libertarianism reject paternalistic policies arguing that individuals necessarily know better than the state what satisfy their preferences[17] (Friedman and Friedman, 1990) and suggesting that bureaucrats would not use laws and regulation to improve agents' welfare but rather to achieve their own personal objectives (Becker, 1976; Buchanan, 1959; Stigler, 1971). Furthermore, some supporters of libertarianism argue that autonomy of choices is a fundamental ingredient of welfare (Wright and Ginsburg, 2012). Therefore, any form of paternalism that limits individual's autonomy should be carefully considered, since the increase in material welfare derived from paternalistic regulation might not compensate the welfare losses connected

[16]According to the context, the term libertarianism could assume different and more specific meanings compared to the very broad definition I use here. For example, some libertarian thinkers advocate a role for the central state limited to a set of basic activities (Nozick, 1974) while others propose to completely replace governmental functions with private alternatives (Friedman, 2008). In this work I label "libertarian" any positions that refuse policy interventions trying to limit internalities - costs that people impose on themselves that they don't internalize (Herrnstein et al., 1993).

[17]This position was first expressed by Mill ([1859]/2010, p.80): "The only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others. His own good, either physical or moral, is not a sufficient warrant. He cannot righfully be compelled to do or forbear because it will be better for him to do so, because it will make him happier, because, in the opinion of others, to do so would be wise, or even right".

from deprivation of autonomy (Rebonato, 2012).

1.4.1. Libertarian Paternalism or "The Real Third Way"

In the wake of progresses in behavioral economics and experimantal psychology, scholars in recent years have proposed a new policymaking movement that on one hand addresses the concerns of libertarian philosophy and on the other hand suggests policy interventions that discourage agents from engaging in non-optimal choices. This new movement has been alternatively labeled "light paternalism" (Loewenstein and Haisley, 2007), "soft paternalism" (Sunstein, 2013), "asymmetric paternalism" (Camerer et al., 2003; Loewenstein et al., 2007) or (as we will refer to) "libertarian paternalism" (Thaler and Sunstein, 2003; Sunstein and Thaler, 2003). According to some authors, libertarian paternalism represents the "real third way" in policy intervention. In the words of Richard Thaler and Cass Sunstein:

> "Libertarian paternalism is a relatively weak, soft, and nonintrusive type of paternalism because choices are not blocked, fenced off, or significantly burdened. If people want to smoke cigarettes, to eat a lot of candy, to choose an unsuitable health care plan, or to fail to save for retirement, libertarian paternalists will not force them to do otherwise — or even make things hard for them. Still, the approach we recommend does count as paternalistic, because private and public choice architects are [...] self-consciously attempting to move people in directions that will make their lives better. They nudge." (Thaler and Sunstein, 2008, pp.5-6).

Libertarian paternalism make use of specific behavioral policies that address systematic and predictable violations of rationality to steer agents' behavior in directions that are self-beneficial without limiting individual autonomy or restricting freedom of choice. In fact, models of decision-making typically assume that people consider only the key features of a decision and calculate costs and benefits of any possible outcome. The intuition at the basis of libertarian paternalism is the same of any behavioral policy: in reality humans

are characterized by limited cognitive abilities and their choices are often influenced by apparently insignificant details (Loewenstein and Haisley, 2007). For example, the order in which food is presented in a cafeteria greatly influences the amount of vegetables consumed as a fraction of unhealthy items (Thaler and Sunstein, 2003) and the savings retirement plan that a company offers as a default tends to be choosen by a large fraction of employees, no matter which is the contribution rate (Choi et al., 2003; Carroll et al., 2009). Therefore, the cafeteria manager and the employer in the examples are "choice architects", since the way they choose to present the situation to customers and employees will affect the aggregate outcomes. Hence, assuming we all agree that on average people save too little for retirement and that unhealthy food causes diseases and reduces life expectancy, the employer could choose a default option that implies adequate retirement savings[18] and the cafeteria manager a food disposition that maximizes healthy food consumptions. On the other hand, those employees that prefer a retirement plan different from the default could freely choose to opt-out and the customers preferring to consume unhealthy food of course have the freedom to pick up what they desire without facing higher prices or other kinds of constraints.
What distinguishes libertarian paternalism from other forms of behavioral policies is that, in steering agents' behavior toward self-interested choices, it refrains from the forms of coercion typical of the classic paternalism and employs instead a "nudging" approach. A nudge could be defined as an aspect of the choice architecture that will have an influence on agents' decisions in a systematic and predictable way without limiting individuals' choice or modifying the economic incentives. Nudging interventions are therefore characterized by not mandating any particular behavior and by the fact that they are cheap to avoid (Thaler and Sunstein, 2008).

[18]See for example the Pension Protection Act enacted by the US government in 2006 with bypartisan support (Beshears and Weller, 2010).

Critiques to Libertarian Paternalism

Beside encountering the enthusiastic interest of many scholars and policymakers, libertarian paternalism raises also critiques coming from different perspectives. In this section I summarize the most common points made against libertarian paternalism[19]. I divide the critiques into two broad categories: those focusing on autonomy and objections concerning welfare and welfarist arguments. I also underline how the welfare critiques naturally shift back the discussion to the fundamental methodological problem of identifying a welfare criterion suitable for behavioral public policy that we discussed in section 1.3.

Autonomy

The concept of autonomy has always attracted considerable attention among social scientists (Feinberg and Feinberg, 1989). Defenders of autonomy consider freedom of choice, not welfare, as the polar star for policymaking. Therefore, according to this idea, the problem is not whether policymakers interventions are effective or not: the mere intrusion by an external authority in individuals' freedom of choice constitutes an impermissible action, no matter what the outcomes would be. Individuals must have the "right to be wrong" in performing their choices (Thaler and Sunstein, 2003, p.241). According to supporters of autonomy of choice, people should be treated with respect by the government and judgements of individuals cannot be overriden without harming liberty (Rebonato, 2012).

Of course, even among supporters of autonomy of choice there are different views about the acceptability of libertarian paternalistic policies. Some scholars argue that autonomy is an important component of welfare, but they

[19]Since the main objective of this work is not to discuss merits and flaws of libertarian paternalism but rather to propose original contributions within the more general framework of behavioral public policy, I limit myself to report the critiques raised. For a detailed reply to any of the objections reported in this work, I re-address the interested reader to Sunstein and Thaler (2003); Thaler and Sunstein (2008); Loewenstein and Haisley (2007).

still recognize that it is not the only one. Therefore, according to this mild position, any paternalistic intervention, even those libertarian and not mandating behaviors, causes a reduction of welfare. Nonetheless, the same scholars recognize that in certain situations the gains from policy interventions more than compensate the welfare loss caused by a reduction of autonomy. Therefore, even if reluctant to give up autonomy of choice, supporters of this mild position might still accept libertarian policies in some specific contexts (Conly, 2012).

Conversely, other scholars hold a more radical position with respect to libertarian paternalism: it harms people's freedom, and freedom is an end not a mean. Therefore, libertarian paternalistic interventions should be rejected as well as any other policy that overrides individuals' judgements (Wright and Ginsburg, 2012). This critique is anti-consequentialist in nature[20] and scholars supporting this position most often depart from economic reasoning on welfare and focus on the violation of a fundamental *right* to choose freely (Sunstein and Thaler, 2003).

Welfare

A second set of critiques against libertarian paternalism focuses on its welfare implications. I present below a series of welfaristic objections raised against libertarian paternalism.

Information. How possibly could a public official know better than myself what makes my life go well? This is a common and powerful critique that defendants of the freedom of choice raise against any form of paternalism and state intervention. The central argument is that people could sometimes make wrong decisions but on the other hand regulators are certainly more likely to err than individuals. In fact, public officials lack information about individuals' preferences and they could fail to correctly interpret what

[20]For a definition of the consequentialist principle see the next section in this introduction.

people's ends are or what are the best means to achieve these ends. Therefore, according to this argument, as long as an action is not harming others' well-being, individuals should be left free to choose whatever they prefer.

Market Solutions. Suppose that consumers care about saving energy and that refrigerators currently in the market are inefficient: companies competing to achieve market shares will start producing energy-saving refrigerators, people will buy them and companies unable to satisfy people's needs will be driven out of the market. If people have heterogeneous preferences regarding the trade-off between price and energy efficiency, companies will produce differentiated goods that satisfy both needs. Of course companies could occasionally fool consumers, but in the long run competition ensures that inefficient companies are driven out of the market.

Moreover, companies could create new products and services that counteract people's self-control problems. For examples, in order to help those people that want to make sure they have enough money for Christmas presents, banks offer special saving accounts: subscribers can deposit money throughout the year but not withdraw them until the month of December (Thaler and Sunstein, 2008). Other companies produce alarm clocks that run away and hide if a person does not get out of bed on time in order to prevent oversleeping. Free markets are more dynamic than public officials in understanding the needs of people. If individuals have self-control problems, companies will create devices that solve them. Paternalistic interventions instead have the negative effect of freezing competition.

Learning by Mistakes. We learn, and improve ourselves, from our own mistakes. Libertarian paternalism deprives people of the most powerful tool to discover what they like: making wrong choices (Klick and Mitchell, 2006). Therefore, libertarian paternalistic interventions prevent people from stepping ahead in the process of self-determination and create a world of infantilized citizens (Sugden, 2009). As a consequence, libertarian paternalism impairs people's welfare by eliminating the process of learning-by-doing.

Heterogeneity People's preferences present a great deal of variation. Individuals can occasionally make mistakes, but a regulator can only propose a standardized solution that, in the best case scenario, accomodates the taste of the majority but decreases the welfare of the others. According to this critique, people should be left free to choose different ends and to pursue these ends with different means (Glaeser, 2006). In fact, people's choices that we consider as errors might instead just reflect heterogeneity in individual preferences. The private sector is in general more able to recognize and satisfy people's variegated needs than public interventions.

Public Choice and the Capture of the Regulator. Private citizens suffer from behavioral biases, but public officials share the same problem. As a consequence, regulations and interventions enacted by the government will reflect the same biases of its officials (Glaeser, 2006). Moreover, even if public officials are trained to overcome behavioral biases, they might promote interventions aimed at reaching their own objectives rather than increasing citizens' well-being. On the same line of arguments, lobbies and interest groups could capture the regulator and push him to adopt public policies that pursue private interests at the expense of social welfare (Tullock et al., 2002). This risk is even more pronounced when we consider soft policies that are not totally transparent and involve risk of subconscious manipulation (Wright and Ginsburg, 2012).

1.5. Behavioral Science and Policymaking

In this section I tackle a controversial, however vital, issue for the development of behavioral public policy: how to increase the interest of policymakers for behavioral research. Indeed, in recent years the behavioral public policy-

making movement stepped ahead both in Europe[21], in the UK[22] and in the US[23]. Nonetheless, I agree with the "manifesto of complains" written by On Amir, Dan Ariely, Alan Cooke and some others among the most prominent behavioral scientists, where the authors state that "The failure of psychology and behavioral sciences more generally to influence public policy is particularly painful and frustrating in light of the success of its sibling, economics, as the basis for policy recommendations" (Amir et al., 2005, p. 444).
In this section, I first analyze some of the possible causes that prevented a full implementation and development of behavioral public policy. With respect to this issue, I argue that the comparative advantage of economics and the main problem for behavioral sciences is that behavioral scholars too often provide loose definition of concepts and engage in normative welfare evaluations that are not grounded in comprehensive theoretical analysis. I then underline the fact that behavioral scholars often provide too vague policy prescriptions. Moreover, I suggest that the technology used to derive policy prescriptions that intend to be immediately implemented requires more field experimentation. Finally, in the next paragraph I report some examples of possible clear and practical solutions suggested by scholars for the fundamental problem of behavioral public policy, the determination of a suitable welfare criterion.

[21]For an overview of the applications of behavioral policies in the EU see the policy report of Van Bavel et al., 2013, available at `http://ec.europa.eu/dgs/health_consumer/information_sources/docs/30092013_jrc_scientific_policy_report_en.pdf`; for an example of applied policy see the Consumer Rights Directive, art. 31.3, that incorporates important behavioral insights.

[22]The Behavioral Insight Team officially established by UK governments in 2010, see `http://www.behaviouralinsights.co.uk/about-us`

[23]The Obama administration officially stated its intention to starte a Behavioral Insight Team under the direction of Cass Sunstein that resemble the one already established in the UK, see `http://www.foxnews.com/politics/interactive/2013/07/30/behavioral-insights-team-document/`.

How Can Behavioral Science Influence Policymakers?

Theoretical Versus Applied Research. The first point it is important to discuss is the relationship between theoretical and applied research for behavioral policymaking. As we have discussed in the previous section, neoclassical economic theory provide an elaborate benchmark of analysis for empirical works, that in turn are able to generate a set of precise policy prescriptions. Behavioral scholars have been someway weak in two respects regarding this point. First, too often empirical findings lack a clear theoretical framework that coherently unify the positive and the normative analysis[24]. The struggles that we documented in section 1.3 of behavioral economists for finding a suitable welfare criterion might appear odd at a first glance. However, this theoretical accuracy reassures policymakers (even those not directly involved or interested by the technical discussion) about the validity of the analytical methodologies applied and would eventually provide a framework to develop precise policy prescriptions. Behavioral scientists should increasing the amount of research on methodological issues, for example focusing on determining standards for measuring preferences using non-choice data.
The second problems concerns the gap between scientific publications in behavioral sciences and translation into policy applications. Indeed, the general principles derived from scientific research too often are not translated in specific policy prescriptions (Amir et al., 2005, p. 447). Behavioral scientists interested in having direct impact on the society should not expect policy-makers to do the extra steps required for deriving from a scientific publication a specific policy prescription, especially in a situation where a comprehensive theoretical framework has not been established yet. Scholars and researcher should instead actively engage in the policymaking process and exploit channels of communication commonly used by policymakers, including publications in non-scientific journals and direct consultancy.

[24]The same concern applied to empirical legal studies in general is expressed by Fischman (2013), that claims a reunification of "is and "ought" within the legal discipline.

Technology. The goal of scientific experimentation is to isolate and identify the causal effects of a specific factor on the theoretical construct. Therefore, experiments often involve the construction of artificial settings that abstract from context-dependent circumstances. Conversely, policies have always to be applied in specific settings, where multiple concurring stimuli and forces act simultaneously. Therefore, to make behavioral researches more directly applicable to policymaking, researchers should sometimes trade-off scientific precision for field experimentations that are robust to specific context situations.

Vaguely Correct or Precisely Wrong? In principle, the precise answer to a policymaker's question regarding the effects of a specific policy should always been "it depends". In fact, scholars provide a substantial body of evidence that situational factors and even apparently minor details play a key role in determining how certain stimuli affect human behavior (see Kahneman, 2003 for an overview). Conversely, policymakers require clear-cut answers. Therefore, behavioral scientists wanting to engage in the policymaking process sometimes face a trade-off between the need to follow rigorous scientific prescriptions and the necessity to provide clear recommendations.

The next paragraph discusses a situation where this trade-off is present, finding a suitable welfare criterion for behavioral public policymaking. Above I discussed how crucial is this issue for welfare analysis and we have seen that scholars did not reach a consensus regarding the standards to adopt. I report possible solutions that attempt to balance the need of scientific rigor, thus offsetting the risks of an indiscriminate abandon of the revealed preferences approach, with the practical problems faced by policymakers that want to apply behavioral public policies.

What Welfare Criterion? Provide an Imperfect but Pragmatic Approach to Policymakers

A welfare criterion that does not truly depart from the basic assumption of the preference-based approach is known as "Informed Decision Utility". This

criterion requires policymakers to ensure that agents are truly informed when they are making their choices. Hence, it suggests to provide warnings against possible decision biases and to facilitate agents' gathering of information about the object of choice. Furthermore, in situations where agents tend to underappreciate the risks or the long-term consequences of certain actions, informed decision utility policies expose and make these consequences salient to agents[25].

One problem with this approach is that policymakers have to engage in value judgements, deciding among the infinite range of situations where information could be improved which ones require policy interventions. Moreover, as we discussed before, information is unlikely to be "neutral": the choice involved might be affected in opposite ways according to the framing of the information provided. Therefore, deciding how to convey the information involves some form of welfare criterion that is not specified. A second limitation of this approach is that it addresses only problems of suboptimal decisions deriving from a lack of attention or information, but does not offer solutions for mistakes deriving from self-control problems. Either naive agents unaware of the behavioral biases leading them to poor decisions or sophisticated individuals that are seeking for solutions of their self-control problems would actually derive little benefits from just being told about the problem without being offered a solution.

Another criterion for the adoption of behavioral policies has been proposed by Camerer et al. (2003). The authors specify the "ideal" conditions being that the policy would help people that behave suboptimally but has no impact on the behavior of the people that already make optimal choices. Hence, default rules or framing alternatives seem to satisfy this criterion, since they may steer inattent people toward advantageous alternatives without imposing any

[25]For example Loewenstein and Haisley (2007) report an existing program that aims at discouraging childbearing by young mothers not ready for it by providing dolls that require constant attention to teenagers at risk for pregnancy.

mandate to others. On the other hand the authors recognize that many policies, while beneficial for biased agents, would impose costs on those who are rationally choosing the optimal outcome. Hence, they propose a "looser but pragmatic" criterion based on cost-benefit analysis: to implement a policy any time its aggregate benefits to behaviorally biased individuals exceed the costs imposed to unbiased agents. While this criterion is useful in shifting the discussion from the abstract concepts of autonomy and freedom to the more concrete measures of benefits and costs (where losses of freedom and autonomy are treated as a cost), nonetheless it does not address the main point of finding a welfare measure that is not preference-based.
Finally, a more comprehensive proposal has been advanced by Loewenstein and Haisley (2007, p.221). The authors argue that behavioral policies should be safely implemented when "welfare judgement tend to be relatively straighforward". In order to identify these situations, they propose a set of sufficient conditions:

- Dominance: there are frequent situations in which people simply "leave money on the table", as in the case of an employee that could contribute to her saving account, benefitting from the employer's match and withdrawing the full deposit the same day without penalty (Choi et al., 2011). Unless we rely on the unrealistic assumption that people show non-monotonic preferences for money, in these situation it is clear that some behavioral bias is the cause of suboptimal decision outcomes. This criterion could be also extended to stochastic dominance. According to stochastic dominance, policy interventions are justified if, in a situation involving an agent's choice under risk, the returns are maximized at any possible level of risk. For example, people including their own stock in their retirement portfolio show a behavior that violates stochastic dominance.

- Clearly Negative Outcomes: sometimes people's decisions generate outcomes that are detrimental under any perspective. For example, many

householders in the US borrow from credit cards at a rate of approximally 18% and at the same time lend money getting a fix return of 6% (Sunstein and Thaler, 2003). In a situation like this, people simply fail to take advantage of an arbitrage opportunity, leaving therefore money on the table. Libertarian policy interventions seem not to require further justifications in similar situations.

- Self-officiating: Obese people, gamblers or drug addicted constantly report they would be better off were they able to modify their behavior regarding food, gambling or drug consumption. In these situations, it seems reasonable to implement libertarian paternalistic policies to help them achieving the desired goals. Loewenstein and Haisley (2007) state this condition specifying they embrace a concept of welfare based on preferences rather than choices. In fact the authors recognize that in certain situations behavioral biases might drive individual choices in directions not reflecting inherent preferences.

2. From Individual to Aggregate Welfare: Policy Analysis and the Choice of the Social Welfare Function[26]

In the last section of the previous chapter, I discussed one fundamental methodological problem specific of behavioral public policymaking, that is finding an appropriate welfare criterion to conduct social policy analysis. In this chapter, I consider a related however different methodological problem, the aggregation of individuals' well-being in a unique measure of social welfare. This problem is not specific of behavioral public policy but, more generally, affects every approach to social policy analysis.

In order to introduce the problem that the social analyst faces when he has to perform a social policy analysis consider an example. Imagine the major of a city that faces the possibility of a traffic plan reorganization. The plan considers opening to traffic a new street in the city center. The new street would cause an increase of daily earnings of Bill's mini-market located there from \$2 to \$3. On the other hand, it will overall increase the number of people shopping in the area instead of driving outside town, and nearby John's supermarket will decrease its daily earnings from \$6 to \$4.5. The reorganization does not affect any other agent in the city and the major in evaluating the plan only cares about the well-being of business and people belonging to his city. Should the major proceed with the plan implementation?

In analyzing the situation, the major could be interested in maximizing the

[26]This chapter is largely based on my paper "When Choosing the Social Welfare Function Really Matters: a Quantitative Analysis", *Rotterdam Institute of Law and Economics (RILE) Working Paper Series* No. 2013/01, that I coauthored with Diogo Gerard. The idea to write this paper was originnaly suggested by Francesco Parisi during a private conversation: I am deeply indebited with professor Parisi for his comments, suggestions and the support received while writing this work. I also thank Emanuela Carbonara, Robert Cooter, Michael Faure, Jonathan Klick, Louis Visscher and also conference and workshop participants at the 2013 European Association of Law and Economics Annual Meeting, the 2013 German Association of Law and Economics Annual Meeting and the Institute of Law and Economics at Erasmus University Rotterdam. The usual disclaimer applies.

sum of his fellow citizens' *wealth*. Therefore, since implementing the plan would reduce total wealth from \$8 to \$7.5, the new plan will be dropped. Alternatively, the major could be interested in increasing *proportional wealth*, that is he would implement the plan only if the percentage increase in wealth of the benefited agent exceeds that of the agent made worse off. In this second case the plan would be implemented, since \$(2*6)=\$12 $<$ \$(3*4.5)=\$13.5. Or again, the major could be interested in maximizing the *sum* of his fellow citizens' *utility*[27]. If he estimates that agents have a utility function of the form f(x)=$\sqrt{x}$, then the plan will be dropped, since \$($\sqrt{2}$+ $\sqrt{6}$) = \$3.86 $>$ \$($\sqrt{3}$+$\sqrt{4.5}$) = \$3.85. Alternatively, it is possible that the major, in light of some extra-welfarist consideration, attaches a greater weight to Bill's achieving and so he will decide to implement the plan. De facto, the initial question does not have a clear-cut answer.

Before proceeding, let us move from our example to a more general framework of analysis. Consider evaluating the welfare-effect of any policy affecting two agents or interest groups, A and B. The policy determines a welfare increase for group A while worsening the wealth of group B. The status quo is $W = (w_a, w_b)$, while the allocation after the policy is implemented is $W' = (w_a+k\epsilon, w_b-\epsilon)$, where k and ϵ are positive numbers. We could interpret the parameter k as the degree of inefficiency implied by the redistribution. For example, when k=1, the system of redistribution is perfectly efficient and resources can be freely transferred between individuals at no cost. Instead, if k<1, we are in a "leaky bucket" situation, whereby redistributing resources implies an overall deadweight loss. While $k\in$[0;1] represents the majority of the situations involving redistributions, nevertheless in some specific cases, it also possible that k>1. For example, we could think of situations where the transfer happens from an unproductive agent (e.g. a rent-seeker) to an agent that creates new wealth (e.g. a start-up). Therefore, every time $k\neq1$

[27]Assuming utility is strictly a function of wealth and interpersonal comparable.

there is a trade-off between equity and efficiency.

The aforementioned difficulty to provide a clear-cut policy evaluation is connected on the one hand to the problems involved in estimating the correct subjects' utility functional form. On the other hand, the result of the social policy evaluation is affected by the value judgments that the policy analyst implicitly assumes. Policy analyst's preferences for redistribution are in fact expressed in the specific method chosen for aggregating individuals' well-being. The analytical tool employed by the social analyst for aggregating individuals' well-being in a unitarian measure is called the social welfare function (SWF onward; I will use this acronym also for the plural cases).

The concept of SWF was introduced by Bergson (1938) and Samuelson (1947) in order to enable formal analysis in the area of welfare economics. Very broadly speaking, a SWF could be defined as a real value function determining social welfare, "the value of which is understood to depend on all the variables that might be considered as affecting welfare" (Bergson, 1938, p.417). Indeed, the analytical representation of aggregate individuals' preferences by means of a SWF allows obtaining a social ordering over alternative possible states of the world. Therefore, the construction of a SWF involves a two-step procedure that follows the process described above: first, making interpersonal comparisons of utility and subsequently aggregating the measures of individual utilities.

Regarding the first step, interpersonal comparisons of utility involves providing a description of individuals' preferences. Economists describe individual preferences indirectly by means of a utility function, an analytical tool that ranks individuals' ordering of alternative states of the world. By defining a utility function, a policy analyst assigns utility indexes to different states of the world in order to reflect their ranking. However, the choice of any particular numerical representation of a utility function that maintains the original ordering is in principle correct, given that utility numbers have no

intrinsic value other than to describe the rank-ordering of the individual[28]. Therefore, a policy analyst may choose different numerical representations of a utility function, which means that ex-ante there is not one most appropriate choice. Hence, it must be noted for the purpose of the present work that this analyst's choice implicitly produces normative statements regarding income distribution. An example presenting the problem graphically is also proposed in Figure 1.

Also regarding the aggregation of individual preferences, value judgments concerning the issue of redistribution are involved in the choice of an aggregation method. In fact, it is possible to choose different functional forms in order to aggregate individual preferences, with each of these functions reflecting a different normative view of what constitutes social welfare. For example, as we discuss in detail in the next section, while social welfare is composed by the sum of individuals' utilities under the utilitarian approach, under the approach usually associated to the philosopher John Rawls (1971), the utility of the least well-off individual determines the social welfare of the society.

The present work aims to enrich the existing literature on social policy analysis concerned with the problem of efficient and fair resources allocation. Throughout the chapter, I revise some fairness criteria among those commonly used in law, economics and political science when performing policy analysis. However, I do not enter into the debate on which is the ethically or philosophically appropriate criterion the policy analyst should embrace. Instead, I provide a quantitative analysis of the relationship among alternative criteria specifications that reflect different distributional preferences. My goal is to quantitatively define the implicit relationships existing between different criteria of fairness that are commonly embraced by the analyst in

[28]For the discussion about cardinal and ordinal utility functions and further references see the Introduction to this thesis.

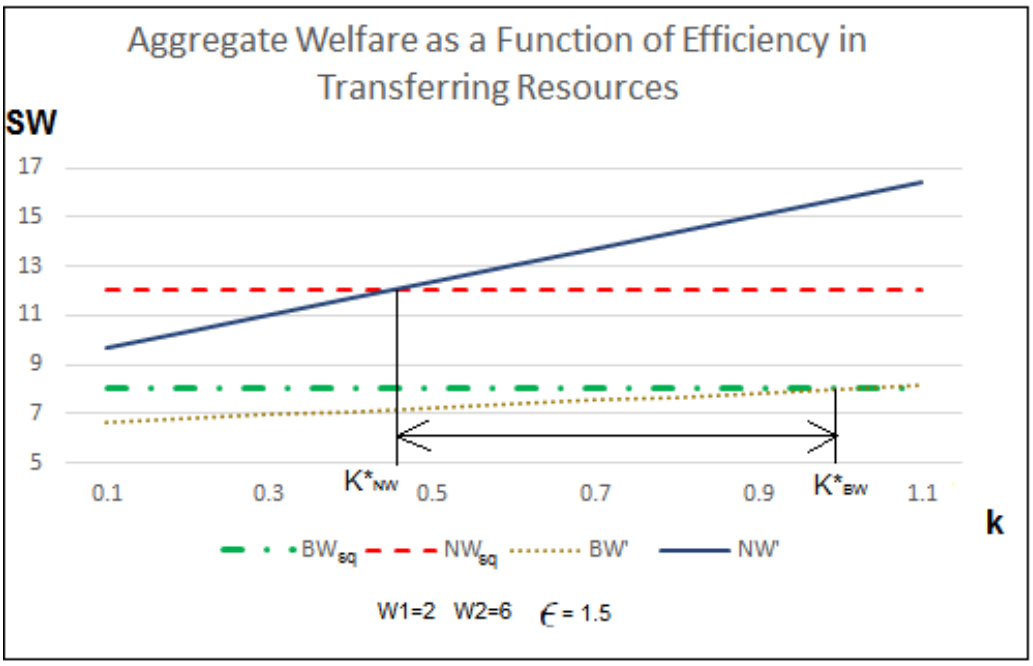


Figure 1: An example of a redistributive policy evaluation that might lead to opposite outcomes according to the social welfare function chosen for the analysis. Given agents' wealth W1 and W2, BW_{sq} and NW_{sq} represent the social welfare of the status quo, calculated respectively by the sum of individuals' wealth and by its multiplication. Conversely, BW' and NW' represent the social welfare after the redistribution of a quantity ϵ of resources from W2 to W1. BW' and NW' are plotted as a function of k, the level of efficiency in transferring resources among agents. NW' is evaluated as a welfare improvement of the status quo for levels of efficiency in transferring resources bigger than k^*_{NW}, while BW' is considered an improvement of the status quo when k>k^*_{BW}. Therefore, if the actual level of efficiency in transferring resources k^* is k^*_{NW}<k<k^*_{BW}, the analyst's choice of the social welfare function will determine the policy evaluation result.

performing social policy evaluations.

A multitude of fairness criteria has been proposed in philosophy and social sciences. In this chapter I restrict the analysis to a selected subset of these criteria. On one hand, I distinguish between individual wealth and utility as indicators of individual well-being. Within the utility specification, I further differentiate between the cases of exponential, polynomial and logarithmic utility functions. Furthermore, I consider different aggregation methods that

correspond to alternative perspectives on the idea of fairness, namely: the sum of individuals' welfare ("Bentham criterion"); the multiplication ("Nash criterion"); and the maximization of the less well-off individual's welfare ("Rawls criterion").

I restrict the analysis to these three notions for the following reasons. First, I want to limit the discussion to cases that are the most prominent and commonly used in law and economics, as well as in welfare economics literature. Moreover, as I will discuss in section 3, precise definitions and axiomatic characterizations have been provided in the context of economic theory for the three criteria considered. Finally, the three cases cover the entire range of possible redistribution considerations, from the fairness-neutrality perspective of utilitarianism to the focus on the least well-off agent of the Rawlsian theory, passing through the intermediate position adopted by Nash. As an additional extension, I further investigate the possibility of generalizing results from a policy evaluation under the two following cases: first, transfer happens strictly from better-off agents to worse-off agents ("Robin-hood condition"); or, second, the transfer happens only from poorer to richer agents and increases total wealth ("Efficiency Increasing condition").

I make the following contributions to the debate regarding the choice of the social welfare function for conducting policy analysis:

1. I formally show that, in general, different combinations of individual well-being evaluation and aggregation methods rank alternative states of the world differently.
2. However, I demonstrate that some apparently distinct combinations always rank alternative states of the world in the same order.
3. Considering a two-agent situation, I derive quantitative conditions necessary to generalize the policy evaluation results obtained implementing a specific combination of individual utility and aggregation method to the other social welfare functional forms considered.
4. I extend the analysis by allowing weight attached to agents (or groups

size) to differ.

The present work is the first of its kind in the fields of law, welfare economics and policy analysis.

The remainder of this chapter is structured as follows. In section 2.2, I provide an overview of the literature in economics and law and economics dealing with resources allocation. In section 2.3 I introduce notation, specify formal definitions and provide a discussion about the three aggregation criteria considered. In section 2.4 I formally show the impossibility of generalizing results obtained with a specific SWF, underlying exceptions and special cases. In section 2.5 I derive the quantitative conditions necessary for the generalization of results in a two-agent scenario. Section 2.6 concludes the chapter, summarizing the results obtained and possible applications. Tables reporting the quantitative results and relative proofs are provided in Appendix A.

2.1. Fairness and Justice, still an Open Debate: Literature Review

Researchers in the social sciences take very different philosophical and ethical positions regarding what constitutes social welfare and the proper distribution of resources. As a consequence, in the branches of welfare economics operating interpersonal comparison of well-being, different SWF forms have been proposed, each reflecting a specific interpretation and value judgment (in the following discussion, I focus on the SWF forms most common in the literature; for a survey considering also less commonly use SWF forms, see Young, 1995 and for a discussion see Foster and Sen, 1997; Ng, 2007). Due to this multiplicity of normative criteria, the assessment of a social policy desirability in practice might depend upon the analyst's choice of the SWF. Aware of the importance of this issue, welfare economics scholars have extensively contributed to the discussion in recent decades. Axiomatic characterizations of the concept of distributive justice and various SWF forms have been proposed, with each of them reflecting a different standard of fairness

(see for example Mirrlees, 1971; Mas-Colell et al., 1995; Sen, 1973; Muller, 1989; Ng, 2000; in the next section, I discuss in detail some of the normative positions advanced by these authors). I am aware that the concepts of "fairness" and "justice", while sometimes difficult to disentangle, nevertheless refer to different notions (for an in-depth discussion, see Rawls, 2001). Despite such differences, scholars in the social sciences sometimes use them interchangeably, following Rawl's argument that a society may in practice only be "just" if it is also "fair" (Rawls, 1958, 1985; however, there are critiques of this position, see for example Knight, 1998 and Sen, 2009). Whereas a discussion of this issue lies beyond the scope of the present work, I strive to be consistent by employing the notion of "fairness" referring to a "perceived appropriateness of the distribution of goods and services". Furthermore, I also specify in the text when contributions that use the word "fairness" with a meaning distinct from mine are reviewed, as for example in Kaplow and Shavell (2009, 2001, 1999).

Indeed, the relationship between the concepts of efficiency and justice also remains an open issue in other fields that apply economic analysis, for example law and economics (Parisi and Rowley, 2005). Posner (1985) contributes to the debate, defending the concept of wealth maximization as a guide for judicial action. According to the concept of wealth maximization, judicial action should promote the activities implying the creation of the highest achievable level of wealth for the society. By contrast, Calabresi (1970, 1985) argues that justice is an end of different order with respect to efficiency, which is only one of the components of this more complex notion. Hence, from his perspective an increase in wealth may not realize a social improvement if disconnected from fairness considerations. Therefore, in Calabresi's view the function of law and economics discipline involves the analysis not of justice itself, but rather of certain ingredients, such as efficiency, concurring to shape the notion of justice.

In a series of papers, Kaplow and Shavell (1999, 2001, 2009) adopt a strictly

welfaristic approach and defend the idea that social decisions must be based exclusively on their effects on individuals' well-being, thus excluding any element of what they call "fairness" from the analysis. However, it should be noticed that what the authors mean by "fairness" substantially differs from the concept we use in the present work. In order to define the notions of fairness, the authors state that "evaluations relying on it are not based exclusively – and sometimes are not dependent at all – on how legal policies affect individuals' well-being" (Kaplow and Shavell, 2009, p.39). Hence, according to the authors' terminology, the concept of fairness includes ideas of justice, natural rights and similar concepts that do not have a direct link with individuals' welfare. However, it must be underlined that by excluding fairness from social decision-making, Kaplow and Shavell do not suggest that economic analysis should avoid making normative statements, particularly regarding the problem of redistributive justice. As the authors makes clear, the argument of their contributions applies independently of which particular position regarding distributive justice has been embraced (Kaplow and Shavell, 2009, Ch.II, pp.15-38). Kaplow and Shavell's works contributed to arouse the discussion among legal and economic scholars concerning what constitutes social welfare, as well as its relationship with the concept of justice (see, among others, Craswell, 2003; Chang, 2000; Dorff, 2002; Fleurbaey et al., 2003; Spector, 2004).

Nonetheless, despite such rich series of qualitative investigations of the concept of fairness, there has been little work in welfare economics, legal disciplines or political science to quantitatively define the relationship between the different SWF specifications. In fact, contributions deriving quantitative results have been produced almost exclusively outside the specific domain of social sciences. For example, the relationship between fairness and efficiency in resource allocation problems has been investigated in engineering applications in communication networks (Luo et al., 2004), air traffic flow management problems (Terrab and Odoni, 1993) or financial applications

and the multi-account optimization problem (see for example Khodadadi et al., 2006). However, such works usually focus on specific situations and thus their results cannot be easily generalized.
Bertsimas et al. (2011) is the only contribution that quantitatively provide insights regarding the relationship between fairness and efficiency in a more general framework. The authors consider different fairness schemes in the context of a resource allocation problem involving multiple agents. Adopting the allocation that maximizes the sum of utilities as a reference point for optimality, they calculate the loss of efficiency implied by switching from an optimal allocation to other more fairness-concerned schemes. Accordingly, Bertimas et al. quantitatively estimate this "price of fairness" in the context of a resources allocation problem, showing how the loss of efficiency associated with fairness-concerned resources allocation schemes variates with the number of players involved in the allocation problem.
Before proceeding with my contribution, in the next section I provide a formal definitions of the concepts I employ.

2.2. Definitions and Research Questions

Following Sen (1970), I define a SWF as a real-valued function that ranks conceivable social states (alternative complete descriptions of the society) from lowest to highest. Inputs of the function include any variables considered to affect the economic welfare of a society. One use of SWF relevant for the present discussion is to represent prospective patterns of collective choice regarding alternative social states.
Formally, a SWF could be defined as follow:

$$\text{SW(W)}: \mathbb{R}^n \longrightarrow \mathbb{R}$$

where $W = (V_1, V_2, V_3, ..., V_n)$ is a vector containing the welfare of each single individual in the population. The only assumption I impose is the function to be weakly increasing in the welfare of each single individual. Formally, if

$W^1 \geq W^2$, then $SW(W^1) \geq SW(W^2)$.

Note that, in principle, this formulation is general enough to accommodate any desired measure of welfare as a maximand. In this work, I consider the cases where it is defined as an individual's wealth or utility, given that these are the most commonly used in the literature.

Regarding the specific form of SWF, the widespread interest in fairness within social sciences has resulted in a multiplicity of principles that have been proposed and applied. However, in practice the criterion used in most policy analysis is restricted to a small number of SWF forms. Specifically, these most frequently applied functions critically diverge in the way in which they embody the trade-off between efficiency and fairness (for an in-depth discussion on this topic see Young, 1995 and Sen and Foster, 1997). I present the three most commonly used SWF forms in the social sciences.

First of all, I introduce the SWF inspired by the work of the British philosopher Jeremy Bentham, labelled "classical utilitarianism". The utilitarian principle has the objective of maximizing the sum of agents' welfare, or, using Bentham's words, "the greatest happiness of the greatest number that is the measure of right and wrong" (Bentham, 1776, Preface, p.ii). Hence, I define a "Bentham" SWF as:

$$SWB(W) = \sum_{i=1}^{N} V_i \tag{1}$$

The problem with any utilitarian solution, also called the Bentham-Edgeworth solution, is the entire absence of fairness considerations. Hence, from an ethical standard, the acceptability of the utilitarian principle has been questioned from more perspectives (see for example Gauthier, 1963; Nagel, 1970; Rawls, 1971).

A different perspective is derived from Nash's studies of bargaining solutions,

the Nash standard of comparison (Nash, 1950)[29]. Under the framework proposed by Nash, a transfer of resources between agents is justified if the percentage of welfare increase of the gainer is greater than the loser's percentage loss. Therefore, a "Nash" SWF can be defined as:

$$SWN(W) = \prod_{i=1}^{N} V_i \tag{2}$$

Finally, I consider a SWF form inspired by the work of the American philosopher Rawls (1971, 1974a,b). According to this theoretical position, the welfare of a society is constituted by the welfare of its least well-off individual. Hence, any welfare-enhancing policy should involve maximizing the welfare of the worse-off agent. Therefore, a "Rawls" SWF could be formally written as:

$$SWR(W) = min_{i=1}^{N} \{V_i\} \tag{3}$$

The three forms of SWF I consider in this chapter could fairly represent the full range of preferences concerning the trade-off between efficiency and distributive fairness. The Bentham SWF implies a full concern for efficiency and no consideration for fairness. In fact, an increase in the total amount of wealth is considered a welfare improvement, no matter if it comes at the expense of the worse-off agent and if it further increases inequality. The opposite position is implied by the Rawls SWF, that is only concerned about fairness and does not consider efficiency. In fact, according to a Rawls SWF the welfare of a society is evaluated by the welfare of its worst-off individual.

[29]Kalai and Smorodinsky (1975) propose and axiomatize an alternative solution to the Nash standard of comparison. The two solutions differ for the set of axioms that they are able to satisfy. In particular, the Nash solution satisfies Pareto optimality, symmetry, affine invariance and independence of irrelevant alternatives. The Kalai-Smorodinsky solution, beside the first three axioms of the Nash solution, satisfies also monotonicity, however it is not able to satisfy the independence of irrelevant alternatives.

Therefore, no social welfare increase is achievable if the poorest agent does not increase his situation. Finally, the Nash SWF assume an intermediate position between Bentham and Rawls. On the one hand, it is concerned about efficiency and in some situations might consider a welfare improvement an activity that further increases the welfare of the better-off agent while it reduces the welfare of the worse-off agent. On the other hand, the Nash SWF implicitly weights more the welfare of the poorest agent, since to consider welfare-improving a policy it requires that the percentage of welfare increase of the gainer is greater than the loser's percentage loss. In fact, holding constant a given amount of resources, the poorer an agent, the greater is the percentage of total wealth represented by this amount.

I now move to another dimension of the problem, presenting the forms of maximand that I consider in this paper. In broad terms, within the literature one most frequently chooses to use either the wealth or some measure of utility as a measure of individual well-being. These possibilities differ in two fundamental aspects that are closely related. First, since marginal utility of wealth is commonly assumed to be strictly decreasing, choosing utility as a maximand is equivalent to stating that individuals are risk averse to some degree. However, the same does not apply to wealth, which implies risk neutrality concerning individuals' welfare. The second point is that, due to decreasing marginal utility, the use of the utility function as a maximand embodies some distributional concerns because one extra unit of wealth is more valuable to a worse-off individual in comparison to someone better-off. The degree to which the utility function embodies risk aversion and distributional concerns hinges on the particular choice of the utility function and its degree of concavity. Therefore, I consider three forms of utility function commonly used: polynomial (constant relative risk aversion), logarithmic (constant relative risk aversion) and exponential (constant absolute risk aversion). To summarize, I consider the following maximands:

- Wealth: $V_i = w_i$, where w_i is the wealth of the individual i

- Polynomial Utility: $V_i = u(w_i) = w_i^{\alpha}$, where $\alpha \in (0, 1)$
- Logarithmic Utility: $V_i = u(w_i) = ln(w_i)$
- Exponential Utility: $V_i = u(w_i) = 1 - e^{-\alpha w_i}$, where $\alpha \in (0, 1)$

By combining these 4 maximands with the three aggregation methods described above, I find the twelve specifications analyzed in this chapter and summarized in Table 1.

Table 1: Social Welfare Function Specifications

Maximand/SWF Form	Bentham	Nash	Rawls
Wealth	$\sum_{i=1}^{N} w_i$	$\Pi_{i=1}^{N} w_i$	$min_{i=1}^{N} \{w_i\}$
Polynomial	$\sum_{i=1}^{N} w_i^{\alpha}$	$\Pi_{i=1}^{N} w_i^{\alpha}$	$min_{i=1}^{N} \{w_i^{\alpha}\}$
Logarithmic	$\sum_{i=1}^{N} ln(w_i)$	$\Pi_{i=1}^{N} ln(w_i)$	$min_{i=1}^{N} \{ln(w_i)\}$
Exponential	$\sum_{i=1}^{N} (1 - e^{-\alpha w_i})$	$\Pi_{i=1}^{N} (1 - e^{-\alpha w_i})$	$min_{i=1}^{N} \{(1 - e^{-\alpha w_i})\}$

In the next section, I formally show that, in general, it is not possible to extend the result of a policy evaluation obtained choosing a particular form of SWF to analysis that employ different SWFs. However, I discuss separately exceptions and particular cases in which the results obtained are robust to generalizations.

2.3. General Results and Special Cases

2.3.1. Generality of Results obtained under a particular SWF: different SWFs rank alternative states of the world in different orders

In this section, I show that any specific form of SWF could potentially rank preferences over states differently to other SWF forms. Hence, an improvement in welfare a reseracher claims when he analyzes a situation employing a specific SWF specification might not be confirmed by other analyses that use different SWF. An exception to this statement is represented by some special cases discussed thereafter.

The intuition behind my way of proceeding is the following. I start by considering the marginal effects of two distinct individuals' welfare on a particular SWF form. I subsequently compare the ratio of these marginal effects with the ratio of the same marginal effects derived assuming a different SWF form. If the ratio of individuals' marginal effects differs among the two SWF specifications, this means it is always possible to reallocate resources within individuals in such a way that the new allocation is welfare-improving for one of the SWFs considered. Specifically, when I subtract one unit of welfare from individual j, the amount of resources I need to give to individual i to compensate the social welfare loss is lower in the SWF that has the lower relative value of j over i. Therefore, an increase in agent i's welfare that restores social welfare to the initial level in the low-interpersonal value SWF is not sufficient to restore initial social welfare under other higher-interpersonal value SWFs.

Let SW^1 and SW^2 be two social welfare functions ordering the social planner's preferences over states of welfare, W, which is a vector of "n" components containing the wealth of each individual.

Proposition 1. *If* $\frac{\frac{\partial SW^1}{\partial w_j}}{\frac{\partial SW^1}{\partial w_i}}$ *and* $\frac{\frac{\partial SW^2}{\partial w_j}}{\frac{\partial SW^2}{\partial w_i}}$ *exist, and for some j,i,* $\frac{\frac{\partial SW^1}{\partial w_j}}{\frac{\partial SW^1}{\partial w_i}} \neq \frac{\frac{\partial SW^2}{\partial w_j}}{\frac{\partial SW^2}{\partial w_i}}$, *then* SW^1 *and* SW^2 *are not equivalent, in the sense that they do not rank all possible allocations in the same order.*

Proof. See Appendix A □

Corollary 1. *Given the same maximand, Bentham, Nash and Rawls SWFs do not necessarily rank alternative states of the world equally.*

Proof. See Appendix A □

Proposition 1 shows that in order to have two distinct SWFs ranking all the possible states of the world equally, the relative value of all pairs of an

individual's wealth must be the same among the two functions at any point in the domain. This implies that a modification of the status quo may or may not be considered welfare improving, depending on the SWF specification selected by the policy analysts.

To conclude, I have shown that the evaluation of a policy derived using a specific SWF does not necessarily hold when performing the analysis with other SWF forms. However, I show in the next paragraph that there are exceptions and particular cases to this proposition.

2.3.2. Exceptions and Special Cases

Now I show that, even though different forms of SWF generally yield distinct preferences over states of the world, there are nevertheless particular cases where generalizations are possible. I proceed by showing that some commonly used SWFs provide equivalent policy evaluation results.

Proposition 2. *The following paired choices of SWF form and maximand always rank alternative states of the world in the same order:*

- *Nash SWF with wealth:* $SWN^{w}(W) = \Pi_{i=1}^{N} w_i$
- *Nash SWF with polynomial utility:* $SWN^{pol}(W) = \Pi_{i=1}^{N} w_i^{\alpha}, \alpha \in (0,1)$
- *Bentham SWF with logarithmic utility:* $SWB^{log}(W) = \sum_{i=1}^{N} lnw_i$

Proof. See Appendix A □

Proposition 2 states an interesting equivalence between some forms of Bentham and Nash aggregation methods paired with different maximands. It is worth noticing this analogy, because Bentham and Nash criteria reflect distinct perspectives regarding the trade-off between equity and efficiency. Indeed, a Bentham aggregation method gives more weight to efficiency than

a Nash one. However, I have shown that the choice of a logarithmic maximand coupled with a Bentham SWF would produce exactly the same results as a Nash-polynomial SWF. Hence, the alternative ethical and philosophical values embodied by these two aggregation methods do not lead to different policy evaluation results when specific individual utility functions that "counteract" such differences are chosen.

As shown in the present and previous subsections, the generalization of results is only possible within a subset of SWF specifications. Therefore, in the following section I derive quantitative results for the possibility of generalizing policy evaluation results across the SWF forms considered.

2.4. Assessing Generality of Policy Analysis Results

From what I have shown in the previous sections, one might be tempted to argue that any policy evaluation result is only valid as long as the ethical and philosophical values embodied in the SWF chosen by the policymaker are considered acceptable. In particular, I have already underlined that choosing a specific SWF in evaluating a redistributive policy implies a normative value judgment regarding the trade-off between inefficiency due to redistribution and fairness. However, despite the aforementioned impossibility of a straightforward generalization, in this section I investigate and compare the robustness of policy results obtained through different SWF. In particular, I want to check whether, all things being equal, there are SWF specifications whose results are more general than others, in the sense that they accommodate a broader class of value judgments.

I consider a two-agent (or interest groups) scenario. One reason for the introduction of this simplified framework is that it is often the case in either theoretical literature or applications that a policy introduction only directly affects two specific interest groups, leaving the other population members unaffected. Moreover, a two-agent scenario allows deriving useful general quantitative insights about the relationship intercurring between alternative SWF specifications, avoiding at the same time analytical complications. I

initially derive quantitative conditions required for extending results without imposing restrictions. I then proceed in later subsections by focusing the analysis on either fairness-improving or efficiency-improving transfers. In the former case, I restrict the attention to the reallocation of resources only benefitting the worst agent at the expense of the best-off. Instead, in the latter case I consider transfers that increase the overall quantity of social wealth at the cost of decreasing equality. As we will see, these restrictions allow for some generalizations of results. Furthermore, they also reflect the majority of real world situations in which a prospected policy is evaluated, given that it would be straightforward to evaluate as desirable a policy that increases societal wealth and at the same time reduces inequality.

I further differentiate the analysis with respect to group size[30]. While I consider interest groups to have the same population size in the next subsections, by contrast, I investigate how the relationship between different SWF changes when group sizes are non-homogeneous in the final subsection. The idea is to consider the effects of a policy change that affects both parties, increasing the welfare of one party while making the other worse-off. The objective is to determine how the conditions necessary to register a societal welfare-improvement vary across different SWFs.

2.4.1. Groups of Homogeneous Size

In this section I consider aggregation methods and maximands summarized in Table 1: three aggregation criteria, Bentham (B), Nash (N) and Rawls (R), combined with four possible individual utility function specifications, utility equal to wealth (W), polynomial utility (P), logarithmic utility (L) and exponential utility (E). Table A.4 in Appendix A reports the minimal level of k necessary to rank the new allocation generated by the redistribution weakly

[30]It is also possible to interpret this situation as assigning different weights to individuals in a two-agent resources allocation problem, a procedure commonly used in extra-welfaristic policy analysis.

preferred to the status quo for any of the combinations examined. While it is not possible to prove general results if we do not impose restrictions, Table A.4 could be useful for conducting a case-by-case analysis. In fact, although results are not very tractable closed forms, it is possible to check whether the results of a policy evaluated as welfare-improving when analyzed with a given SWF are confirmed under other SWF specifications.

Let me provide an example in order to clarify the procedure. Assume for instance that a redistributive policy analysis is conducted using a Bentham-exponential (B-E) maximand. The welfare of the groups in the status quo is $W_{A,B} = (10, 6)$, the amount eventually lost by B is $\epsilon = 2$, the gain of A is $k\epsilon$ and the exponential discount factor is α=0.3. Hence, given the status quo welfares and the parameter ϵ, it is sufficient to plug these values into table A.4 to derive the minimum level of the inefficiency parameter k that makes the policy desirable under B-E. In our example, the policy is welfare-improving under B-E if $k \geq 0.474$. Given this result, it is possible to check which SWFs, holding the parameters value constant, require a lower level of k in order to also consider the policy welfare-improving. In this example, it turns out that N-E and the R specifications require a lower level of k and thus derive a higher level of social welfare from the policy implementation. In fact, a smaller value of k means that, given the welfare loss borne by group B, the welfare increase of group A required in order to increase the overall social welfare vis-à-vis the status quo is lower than that required by the original B-E specification. On the other hand, all the other SWF specifications in the example require a higher minimal level of k, and hence according to these SWFs the policy would be considered welfare-reducing.

It should be noted that Bentham-Logarithm, Nash-Wealth and Nash-Polynomial share the same k. In fact, as shown in Proposition 2, these SWFs rank alternative states of the world in the same way. Another interesting result is that Bentham-Wealth only requires the net difference in wealth to be positive. Finally, it is worth underlining that Bentham-Wealth, Bentham-Polynomial

and Bentham-Logarithm (and Nash-Wealth and Nash-Polynomial) are not sensitive to the unit of measure of wealth. In fact, if we proportionally increase the wealth of both groups and ϵ, k does not change. This happens because, in the former cases, multiplying the wealth of each individual in the population by a constant is equivalent to a monotonic transformation in these SWF, thus preserving the ranking over all possible states. However, the same is not true for Bentham-Exponential, Nash-Logarithm and Nash-Exponential.

2.4.2. Groups of Homogeneous Size and Restrictive Conditions on Transfers Now consider a policy redistribution where transfers are subject to specific conditions. First of all, consider the situation in which transfers of resources are only possible from the best-off to the worst-off group. I name this kind of transfers "Robin-hood" (RH). Furthermore, I also consider the situation in which transfers strictly occur from the less wealthy to the wealthier group and increase total wealth. I label these transfers "Efficiency-increasing" (EI). The intuition is that a policy implementing EI transfers would further increase inequality, but could nonetheless be desirable from a social perspective if the wealthiest group's gains more than compensate the losses borne by the poorest group[31]. Under the condition that transfers are either RH or EI, I derive the following results:

Proposition 3. *After a RH or EI transfer, if the social state alternative to the status quo is preferred according to a certain SWF, it is possible to precisely define the set of SWF that always produces the same result. Specifically:*

- *under a RH transfer, B-W* $\implies$ *B-P* $\implies$ *B-L* $\equiv$ *N-W* $\equiv$ *N-P* $\implies$ *N-L* $\implies$ *R; also B-E* $\implies$ *N-E* $\implies$ *R; and also B-W* $\implies$ *B-E*

[31] I do not include in the analysis transfers that reduce inequality in situations where $k \geq 1$, for the obvious reason that such a transfer would always produce a Pareto improvement and thus would be desirable according to any SWF specification.

$\implies$ *N-E* $\implies$ *R.*

- *under an EI transfer, N-L* $\implies$ *N-P* $\equiv$ *N-W* $\equiv$ *B-L* $\implies$ *B-P* $\implies$ *B-W; and also N-E* $\implies$ *B-E* $\implies$ *B-W.*

Proof. See appendix B. □

Proposition 3 shows that some policy evaluation results are more robust than others and are able to accommodate a broader class of subjective ethical values. For example, consider a RH transfer that is desirable if evaluated assuming a N-P SWF. Given the results presented, one can be sure that the new social state after redistribution would also be preferred by N-W, B-L, N-L and all the R combinations, independently from parties' wealth in the status quo, transfer levels and deadweight losses associated to inefficiency in transferring resources. However, the same is not true if we consider for example a B-P or a B-W SWF. Tables 2 and 3 provide a summary of Proposition 3 results.

Table 2: Generality of Results under RH Transfers

Maximand/SWF Form	Bentham	Nash	Rawls
Wealth	$\{\forall B; \forall N; \forall R\}$	$\{B\text{-}L;\ N\text{-}P;\ N\text{-}L;\ \forall R\}$	$\{\forall R\}$
Polynomial	$\{B\text{-}L;\ N\text{-}W;\ N\text{-}P;\ N\text{-}L;\ \forall R\}$	$\{B\text{-}L;\ N\text{-}W;\ N\text{-}L;\ \forall R\}$	$\{\forall R\}$
Logarithmic	$\{N\text{-}W;\ N\text{-}P;\ N\text{-}L;\ \forall R\}$	$\{\forall R\}$	$\{\forall R\}$
Exponential	$\{N\text{-}E;\ \forall R\}$	$\{\forall R\}$	$\{\forall R\}$

Proposition 3 states that it is not possible to generalize results obtained by, for example, an R-L SWF outside the Rawls aggregation method, even imposing some restrictions on transfers. Hence, any result obtained assuming R-L holds as long as the value judgements implied by the choice of this specific SWF are considered acceptable. By contrast, the result is not necessarily the same if other subjective values are assumed for the analysis. The intuition behind this fact is that, in the case of RH transfers, some redistribution

Table 3: Generality of Results under EI Transfers

Maximand/SWF Form	Bentham	Nash	Rawls
Wealth	$\emptyset$	$\{B\text{-}W;\ B\text{-}P;\ B\text{-}L;\ N\text{-}P\}$	EI¬R
Polynomial	$\{B\text{-}W\}$	$\{B\text{-}W;\ B\text{-}P;\ B\text{-}L;\ N\text{-}W\}$	EI¬R
Logarithmic	$\{B\text{-}W;\ B\text{-}P\}$	$\{B\text{-}W;\ B\text{-}P;\ B\text{-}L;\ N\text{-}W;\ N\text{-}P\}$	EI¬R
Exponential	$\{B\text{-}W\}$	$\{B\text{-}W;\ B\text{-}E\}$	EI¬R

policies are desirable for high inequality-averse SWF specifications, even if the inefficiency deriving from the transfer of resources is huge. Conversely, SWFs reflecting higher concerns for efficiency might evaluate the gains deriving from equality as insufficient to compensate the deadweight loss associated with redistribution. Hence, if a policy that reduces inequality is considered desirable when evaluated assuming highly efficiency-concerned SWFs, it is also likely to be supported under a more fairness-concerned SWF. However, the opposite might not be true as well.

On the other hand, a policy could aim at increasing overall efficiency, even if its implementation would further increase inequality. In an EI transfer scenario, a highly equity-concerned SWF would require higher efficiency gains compared to an efficiency-concerned one. Hence, a higher level of k is required from the former SWF in order to be considered welfare-improving. Finally, we should note from Table 3 that an EI transfer by definition excludes the possibility of any Rawls improvement. In fact, Rawls SWFs are only concerned with equality and, according to the ethical position they reflect, no efficiency gain could compensate an increase in inequality.

2.4.3. Groups of Non-homogeneous Size

I now consider a policy redistribution that affects two interest groups whose number of members differs. Similarly, we could think about a two-agent situation where the analyst assigns different weights to individuals. Let N be the total number of individuals in the two groups. Individuals 1,...,j

belong to group A and j+1,...,N belong to group B. I assume that wealth is homogenous among group members, hence $w_1 = w_2 = ... = w_j$ and $w_{j+1} = w_{j+2} = ... = w_N$. Let $W = (w_1, w_2, ..., w_j, w_{j+1}, ..., w_N)$ be the status quo and $W' = (w_1 + k\epsilon, w_2 + k\epsilon, ..., w_j + k\epsilon, w_{j+1} - \epsilon, ..., w_N - \epsilon)$ be the new allocation where individuals belonging to group A increase their wealth by $k\epsilon$ each, and group B individuals have their wealth reduced by ϵ.

Table A.5 in Appendix A shows the minimal level of k that makes the new allocation weakly superior to the status quo according to each SWF considered. To present the results obtained in a more intuitive way I provide a graphical example. Figure 2 compares the minimal k required by an RH transfer between groups of homogeneous size against the situation where the wealthier group has a size equal to $\frac{3}{4}$ of the other. In particular, we could see that the minimal efficiency level required in the latter situation is lower than that in the former. This directly relates to the fact that a larger fraction of the agents involved benefit from the policy compared to the homogeneous group-size situation. Therefore, even lower level of efficiency in transfering resources might result sufficient to increase social welfare. Furthermore, we should notice that the variation in minimal k between homogeneous and non-homogeneous cases differs among SWF specifications. This is due to the fact that the group size variable introduces an element of non-linearity in some of the equations determining k, while it only results in linear transformations for other specifications.

2.5. Conclusions of Chapter 2

In this chapter, I perform a quantitative analysis of the possibility to generalize policy results obtained implementing a specific social welfare function. I consider common combinations of aggregation methods (Bentham, Nash and Rawls) and utility functions, formally showing that different social welfare functions rank alternative states of the world in different orders, except in a well-defined subset of particular cases. Moreover, adopting a scenario

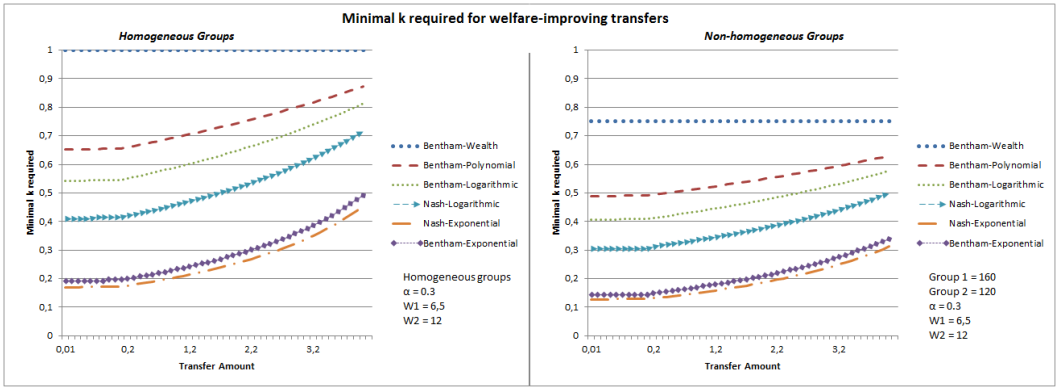


Figure 2: An example of different group-size effects on the evaluation of an RH transfer. The graph on the left has homogeneous groups, while the one on the right involves the wealthier group size being $\frac{3}{4}$ of the less wealthy group. All things being equal, increasing the size of the benefitted group reduces the efficiency level required for evaluating the transfer as welfare-improving.

with two interest groups, I define the quantitative conditions under which a policy is considered welfare-improving for any social welfare function considered. These conditions allow a case-by-case assessment of the generality characterizing the policy evaluation result.

I subsequently proceed by imposing the restrictive conditions that transfers have to be either strictly equity-improving or efficiency-improving.Under these two scenarios, I considered the possibility to generalize the result of a policy evaluation. For each SWF in each scenario, I derived the subset of SWFs that always produce the same policy evaluation result. I show that the possibility of generalizing policy evaluation results crucially depends on the degree of inequality implicitly accepted by a social welfare function. Finally, I repeat the same analysis and derive the same quantitative relationships allowing the two groups of interest considered to have different sizes.

This chapter aims to further enrich the discussion regarding the choice of the most appropriate social welfare function in policy analysis. The results derived in this chapter are important in two respects. First, they have academic relevance for scholars and researchers that employ formal social welfare anal-

ysis. In fact, these scholars could easily verify to what degree the predictions derived in the analyses conducted are sensitive to the specific way they choose to model the situation. Second, findings of this chapter are important for decisionmakers that have the final word regarding policies implementation. Indeed, they could identify to what degree the results obtained in a given policy analysis are sensitive to the specific value judgements adopted by the policy analyst. Threfore, the findings presented in this chapter strengthen the methodology of social policy analysis by clarifying, in a systematic way, the possibility of generalizing results obtained assuming specific normative value judgements.

Finally, it is worth noticing that the applicability of the results derived in this chapter is not limited to the area of behavioral public policy. Indeed, they are relevant also for social policy evaluations grounded in traditional welfarist economics. Moreover, the results are also valid in extra-welfarist analyses of social policy, as for example in cost effectivennes analysis in healthcare. As we have discussed in the previous chapter, the extra-welfarist approaches do not have theoretical foundations in the welfaristic economic tradition nor necessary employ agents' utility as a relevant outcome. Nevertheless, the policy analyst could consider measures of individual well-being other than utility or unequaly weight relevant outcomes and still apply results presented here, as long as the analyst's interest is comparing decision outcomes in term of their monetary cost per unit of effectiveness.

3. Shaping Tax Norms Through Lotteries[32]

The focus of chapter 2 was on the choice of the social welfare function, a methodological problem that characterizes behavioral public policymaking as well as any other approach to social policy evaluation. Instead, in this chapter and in the next one I get to the hearth of libertarian paternalism and I discuss two original policy interventions that exploit nonstandard behavioral regularities in order to achieve welfare improvements. I investigate throughout this chapter a zero cost policy based on rewards aimed at combatting value added tax (VAT) and retail sales tax (RST) evasion. This policy has been applied in some countries and the empirical evidence suggests that it is quite effective. However, according to the theoretical predictions derived from models of standard decision-making, the policy could not increase net tax revenue. In fact, these models predict that the government has to incur costs for the provisions of the rewards higher than the increase in tax revenue collected. So far no explanation has been provided to this puzzling evidence. My goal in this chapter is to propose a model that is able to explain the empirical evidence relative to the policy success, in order to derive theoretical predictions regarding the results of an eventual policy implementation. I present a model based on non-expected utility that is able to explain the policy success. Moreover, I show by means of a calibration exercise that it is

[32]The core sections of this chapter are mostly based on my article "Shaping Tax Norms Through Lotteries", *RILE working paper series*, 2012/02. The introduction, part of the literature review and the sections where policy implications are discussed are mostly based on my article with Sigrid Hemels "Do You Want a Receipt? Combatting VAT and RST Evasion with Lottery Tickets", *Intertax: international tax review*, 2013, 41(8), pp. 430–443. I am grateful to Emanuela Carbonara, Marco Casari, Michael Faure, David Gamage, Andrea Geraci, Jonathan Klick, Francesco Parisi, Matthew Rabin, Louis Visscher and participants to the 2013 European Association of Law and Economics conference, the 2013 Italian Association of Law and Economics conference, the IX Young Economists' Conference on Social Economics at University of Bologna, the VI IMPRS Uncertainty Topics Workshop at Erasmus University Rotterdam and seminars at University of Bologna, Hamburg University, Erasmus University Rotterdam and University Paris II for their support and helpful suggestions. The usual disclaimer applies.

possible to state the conditions necessary to predict a successful implementation of the policy.
The findings of this chapter have relevant implications for policymakers and decisionmakers that are considering the application of this policy and want to predict the likelihood of its successfull implementation.

3.1. Invoices and Indirect Tax Evasion

'When I asked the decorator how much it would cost to paint my house, his answer was: "Do you want a receipt?".' This conversation, overheard during a Dutch birthday party, is an everyday example of an attempt to evade value added tax (VAT)[33]. The decorator would probably ask for a lower fee for painting the house without an invoice as, in that case, he would not charge VAT. An invoice enables tax authorities to carry out controls. Invoices are, therefore, very important in preventing tax evasion and the illegal non-payment or under-payment of taxes. Most VAT and retail sales tax (RST) systems, therefore, include the obligation to issue an invoice[34]. However, this obligation is not always enough to ensure that invoices are actually issued, even if it is accompanied by sanctions in case of non-compliance.
In addition to imposing sanctions on businesses that do not issue invoices, governments can give an incentive to customers to request an invoice and thus obliging suppliers to comply. In this chapter I discuss a specific kind of reward complementary to sanctions and audits to combat evasion of RST and VAT: turning the invoice into a lottery ticket. In the discussion I use empirical Law and Economics research as this research field can give insight into the effectiveness and efficiency of such a compliance strategy. This chapter does

[33]In a column in a Dutch newspaper a similar conversation was published: ' "Do you need the receipt?" Everybody who hires odd-job companies knows this question. As I have just bought my own house, I was not that experienced and asked "The receipt... eh... well... why?" The decorator answered "Well, without one I can give you a good price".' Christiaan Weijts, Fraudeursdromen, NRC Handelsblad 23 October 2012.

[34]For example, article 220 of the Council Directive 2006/112/EC of 28 November 2006 on the common system of value added tax as amended later (hereinafter: VAT-Directive).

not aim to discuss the whole issue of tax evasion and tax compliance: I focus on one specific strategy which is applied to increase RST and VAT compliance. For a general discussion on tax compliance I refer to the vast literature on this topic: Andreoni, Erard and Feinstein even speak of a 'tide of research on tax compliance' (Andreoni et al., 1998).

The structure of this chapter will be as follows. In section 3.2 I discuss the traditional way governments combat VAT and RST evasion, the alternative approach of providing incentives instead of sanctions and engaging consumers as 'unpaid auditors' in enforcing VAT and RST compliance by requiring businesses to issue invoices. Section 3.3 discusses why consumers in certain societies will not ask for invoices to combat tax evasion by comparing this to contributions to public goods, Section 3.4 discusses how consumers can be given an incentive to require an invoice, reports the results of the implementation of lottery ticket invoices in China and discusses the explanation Wan (2010) gave for the success of this policy. As I am not convinced by this explanation, I develop an alternative explanation in section 3.5, proposing a model which can enable governments to decide on introducing lottery tickets or not. Section 3.6 discusses the model implications. Furthermore, in section 3.7 and 3.8 I discuss respectively the possible unintended side-effects and some long-term benefits of this policy. The conclusion in section 3.9 summarizes results.

3.2. Combating evasion of VAT and RST: Literature Review

Slemrod (2007) noted that no government can announce a tax system and then rely on taxpayers' sense of duty to remit what is owed. Andreoni et al. (1998) observed that the problem of tax compliance is as old as taxes themselves. Webley et al. (2006) state that VAT evasion is widespread and involves significant revenue losses. Evasion of VAT and RST is not only a problem in developing countries or in countries in the south of Europe, but in north-

ern European countries as well. In a report of May 2013[35] the European Commission gave an overview of the Actual VAT revenue in 2010 as percentage of theoretical revenue at standard rates. The Commission concluded that Member States are only collecting around one half of the VAT revenue available to them [36]. In the December 2012 Action Plan of the European Commission, VAT fraud and evasion was identified as an important field in which action was necessary. The European Commission shared this view with the EU Member States: "Member States also emphasized the need to adopt quickly the pending proposals in the Council and to pay particular attention to the fight against VAT fraud and evasion."[37] Such evasion not only erodes the income of governments; it also undermines the principles on which government expenditure is shared by citizens of a country and, as a consequence, the division of the tax burden(Slemrod, 2007). Tax evaders are free riders: they benefit from government expenditures without contributing their share to the government income. If nobody would pay VAT or RST, everybody would be worse off as the government would not be able to meet its expenses. If a society is of the opinion that government expenses

[35]European Commission. Combating tax fraud and evasion. Commission contribution to the European Council of 22 May 2013, p. 8.

[36]Several reports have been published on this so called 'VAT gap', the difference between the theoretical amount of VAT that should be due and actual VAT receipts, for example the report of Reckon LLP of 21 September 2009 for the European Commission `http://ec.europa.eu/taxation_customs/resources/documents/taxation/tax_cooperation/combating_tax_fraud/reckon_report_sep2009.pdf`, which gives an overview of the VAT gaps in EU Member States in 2006 and Eurostat/European Commission Taxation trends in the European Union, 2013, p. 31 `http://ec.europa.eu/taxation_customs/resources/documents/taxation/gen_info/economic_analysis/tax_structures/2013/report.pdf`. In this report it was concluded that ten Member States collect less than 50 % of the theoretical amounts, another thirteen countries collect between 50 and 60 % and for only four countries - Bulgaria, Estonia, Cyprus and Luxemburg - the VRR is above 60 %.

[37]Communication from the Commission to the European Parliament and the Council, An Action Plan to strengthen the fight against tax fraud and tax evasion, Brussels, 6 December 2012, COM(2012) 722 final, p.3, `http://ec.europa.eu/taxation_customs/resources/documents/taxation/tax_fraud_evasion/com_2012_722_en.pdf`.

are too high, this should be resolved by reducing government spending in a democratic way, not by the decisions of individuals not to pay their share of democratically set taxes. Tax evasion is therefore undesirable both from an economic and a legal (fairness) point of view. It, therefore, makes sense that governments try to combat the evasion of all taxes, including VAT and RST.

3.2.1. Traditional methods: sanctions on tax evaders

Developing effective policies that promote tax compliance and combat tax evasion is a challenging task for authorities and policymakers. In the words of Andreoni, Erard and Feinstein: "How can an authority – with imperfect ability to monitor - design a taxation, audit, and punishment scheme to meet its revenue objectives?"(Andreoni et al., 1998). Academic research in the field of law and economics can give useful insights into this problem. Traditionally, contributions in law and economics focus on monitoring and sanctions to achieve compliance[38]. Economic models predict that higher penalties and audit probabilities discourage non-compliance, the evidence suggests that higher audit probabilities probably have more impact than higher penalties (Webley et al., 2006). However, Andreoni, Erard and Feinstein (1998) observed that econometric results suggest that the use of the 'stick' to enforce compliance with tax laws may not have any long-run impact. Tax legislation focuses on sanctions as well, such as fines for businesses that do not pay the VAT due on their services and supplies. These traditional methods of combating tax evasion are based on deterrence, the use of sanctions and punishment as a threat to deter taxpayers from offending. However, enforcing the payment of indirect tax through deterrence methods can be costly for the government. Auditing businesses and imposing fines requires that the tax authorities have the means and sufficiently well equipped employees to

[38]Seminal contributions include Allingham and Sandmo (1972) and Yitzhaki (1974). For a survey of the Economics and Law Economics literature on tax evasion see Andreoni et al. (1998).

perform these tasks. Indirect tax payments are based on the financial records of transactions. To establish whether supplies and services have been performed 'outside the books', the tax authorities have to do further research. Due to the information asymmetry between taxpayers (in economic terms: private agents) and the government, a revenue-maximizer taxpayer could be tempted to under report the tax amounts due unless a costly system of monitoring and sanctioning is in place. Sanctions are only effective if they pose a sufficient threat to deter taxpayers from tax evasion.

Efficiency reasoning would lead to setting the sanction at such a level that the marginal cost to the government of monitoring and sanctioning taxpayers would equal the marginal benefit of preventing tax evasion. Theoretically, Becker (1968b) suggests that increasing sanctions would reduce tax evasion. In fact, for a given probability of being detected, the expected profits from evasion are a decreasing function of the level of sanctions. However, there are practical arguments in favor of imposing a sanction ceiling, for example the necessity of preserving the marginal deterrence effect of sanctions and the credibility of the threat made by the sanctioning authority. If strong sanctions are combined with a low risk of tax fraud being discovered and of miscreants actually being fined, these will not be very successful in combating tax fraud. Hence, given the practical impossibility of raising the sanctions level over reasonable thresholds - a death penalty for tax fraud, would, for example, not be accepted in most democratic societies -, we could expect that high monitoring costs will be associated with high levels of tax evasion. Moreover, political constraints could prevent the implementation of sanctions. A legislator interested in maximizing his chances of being re-elected could be "captured" by interest groups benefitting from tax evasion and reduce the chances of effective policies being adopted to combat tax evasion (Stigler, 1971). An example seems to be the failure (unwillingness?) of previous Greek governments to act on the so-called "Lagarde list" of Greeks with

overseas bank accounts[39]. Finally, in specific segments of the population tax evasion could be perceived as a morally justified behavior and pro-tax evasion social norms could develop[40]. The Dutch decorator apparently thought it very normal to offer to do a job with or without VAT. Tax evasion is so deeply rooted in some cultures that it could be considered endemic. For example, during the first half of 2012, in 38 % of the tax audits in Italy (with peaks of over 50 % in some provinces in the south) the issuance of invoices was found to be irregular[41]. These data are confirmed by a recent field experiment run on bakeries in Milan (Battiston and Gamba, 2013). Within a time span of 12 minutes, two customers bought a loaf of bread in 108 bakeries. Only 73 (68 %) bakeries were fully compliant and gave a receipt to both customers. This experiment was performed after much publicity was given to tax audits in shops in several towns, including Milan, and a strong awareness campaign in the mass media. Apparently these campaigns were not enough to completely change the attitude towards the issuing of invoices. In such situations, any coercive intervention by an external authority could be perceived as a violation of the established norm by the targeted population and could produce countervailing effects (Carbonara et al., 2012). Indeed, empirical evidence suggests that, irrespective of the legal and socio-economic context and the effort put into combating indirect tax evasion, it is still a widespread problem (Cowell, 1990; Slemrod, 2007).

3.2.2. Stick and carrot?

The best way to reduce tax evasion would probably be to audit each and every tax payer. However, given the limited means of governments, this is not possible. Even though the traditional methods of deterrence could only mitigate the tax evasion problem, the tax compliance literature has traditionally

[39]L. Thomas. In Greece, Taking Aim At Wealthy Tax Dodgers *New York Times* 11 November 2012.

[40]See section 3.4 for a detailed discussion on this point.

[41]La Repubblica, 31 July 2012.

been skeptical about the possibility of implementing alternative policies (for a discussion of this point see Feld et al., 2006). Nonetheless, some researchers have investigated the effect of implementing reward mechanisms instead of sanctions. Falkinger and Walther (1991) show that a mix of sanctions and rewards would outperform a system with sanctions only without increasing expenditure for the government. Experimental Economics literature has also investigated the effect of rewards compared to sanctions in achieving compliance. For example, Torgler (2003) found in a field experiment among Costa Rican taxpayers that a monetary reward is the most effective way of increasing compliance. In the report of May 2013 on combating tax fraud and evasion, the European Commission also recommended the use of both sanctions and rewards to reduce the size of the shadow economy when it gave the following examples of measures to combat tax evasion: criminalizing the purchaser of undeclared work (sanction) and the use of monetary incentives to declare (reward)[42].

Other research outside the traditional tax policy literature seems to confirm the positive effects of rewards on motivating desired behavior. Both social psychology (Nuttin and Greenwald, 1968; Molm, 1994) and neuroscience (Gray, 1981; Larsen and Ketelaar, 1991) researchers have emphasized the role and effectiveness of rewards in achieving individuals' compliance. In particular, it seems that punishments and rewards have asymmetrical effects on human behavior (Sims, 1980), hence making it possible to reinforce compliance through a combination of the two methods.

However, simply rewarding businesses that comply with their tax obligations seems a bit odd from a legal point of view. The question is, therefore, whether it could be a solution to engage a third party who does not have a legal obligation regarding the tax: the customer in the transaction over which the VAT or RST is due.

[42]European Commission. Combating tax fraud and evasion. Commission contribution to the European Council of 22 May 2013, p. 3.

3.3. Combating Evasion by Engaging Customers: Importance of the Invoice and the Public Goods Trap

In many countries, the invoice is proof of the existence of a taxable transaction. Furthermore, it contains information on the amount of tax due. Once a company has issued the invoice, it becomes difficult, if not impossible, to hide information on the supply and RST or VAT due. Hence, a key strategy adopted by businesses in evading RST or VAT is not to issue an invoice. If customers demand an invoice, this kind of tax evasion is made more difficult. Customers, in a way, act as unpaid auditors for the state, enforcing compliance. In a VAT system, other businesses will ask for such an invoice, as this is necessary for reclaiming the VAT they have paid. However, asking for a receipt has virtually no benefits for individuals who are not taxable for VAT and RST. In fact, as will be discussed in more detail in the next paragraph, without any specific policy intervention, customers not only do not receive benefits, they could also face high social and moral costs when asking for an invoice if it is the social rule not to ask for a receipt.

In economic terms, from the perspective of a consumer, asking for an invoice and thus preventing tax evasion can be compared with contributing to a public good. A public good has two characteristics: it is hard to exclude any person from benefitting from the good or the service even if this person does not pay for it (non-excludability) and the consumption of the good or the service does not prevent the consumption of it by others (non-rivalry). Common examples of such goods are the army and dikes. The non-excludability characteristic of these goods implies that it may be hard to get some individuals to voluntarily pay an adequate share of the costs of a public good, because they cannot be excluded from benefitting from it: the so called free rider problem. Therefore, absent external interventions, the free rider problem would lead to an under provision of the public good. In this section I will analyze whether having to request an invoice could be considered symmetrical to a public goods situation. Economic theory predicts that, because

of the free rider effect, the supply of public goods will be at an inefficient level, below the social optimum. Hence, if the proposed parallel is correct, the enforcement of invoice issuance by customers remains suboptimal if the government does not provide incentives (for a survey on experimental results in public goods games see Ledyard, 1995).

In order to clarify the concept, consider the situation in which a consumer has to claim an invoice from a fraudulent seller. For our purposes, think of the buyer as a potential contributor to a specific public good, namely enforcement of tax payments. The rational buyer evaluates the private costs and benefits of asking for the invoice. For any transaction, the private benefit the individual buyer derives from asking for a receipt is almost zero. The customer hardly benefits himself from the tax the seller pays to the government. In economic terms: the benefit is not fully internalized by the customer. Instead, it is shared with the rest of the population. This is a consequence of the fact that goods financed through taxation are often public in nature and, by definition, non-excludible. The individual buyer and his fellow citizens share the benefit deriving from the tax paid in any transaction even if the latter are not directly involved in the specific transaction.

On the other hand, not asking for an invoice has an economic benefit if the customer can bargain for a discount as compensation for not obtaining a receipt, basically sharing the profit deriving from the tax evasion with the seller. Moreover, even in situations where bargaining is not feasible[43], scholars report evidence of the existence of moral, ethical and social costs

[43]It is often impossible or unprofitable to have a bargaining solution. For example, in transactions involving small amounts of money (such as the loaf of bread in the Milan bakery experiment discussed above), the discount would be negligible or the opportunity cost of the time invested in bargaining would be higher than the discount itself. Moreover, in situations in which face-to-face bargaining is not feasible (e.g. other customers present in the shop, a crowded café, etc.) reputational concerns could prevent a customer from bargaining. Finally, in several countries, such as Japan, bargaining over prices is unusual and considered impolite, so customers would simply reject this approach; on this point see Berton (1998).

for buyers who ask sellers to comply with fiscal norms. McGee (2012) has collected two decades of scholars' contributions on the ethical aspects of tax evasion. His book discusses philosophical and religious determinants of tax evasion, explaining the formation of pro-tax evasion behavioral norms. The author argues that, if the social norm is positive towards tax evasion, individuals wanting to break these norms will face costs. Chang and Lai (2004) proposed a model incorporating social norms into a collaborative tax evasion agreement between a seller and his customer. They found that this collusive practice tends to intensify the tax evasion problem and reduces the effectiveness of tax enforcement. Kirchler (2007) also analyses the behavioral aspects of tax compliance and evasion, focusing on the psychological reasons that lead to customers colluding and accepting tax evasion.

The research mentioned above suggests that in some cultures and societies costs are associated with not complying with the established norms favoring VAT and RST evasion. While the consumer bears the personal costs and sometimes misses the opportunity of a discount in expressly requesting an invoice, he basically gets no benefit from this enforcing operation. Even though requesting an invoice would be optimal from a social point of view, in the above mentioned social contexts free riding on the associated costs remains the individual dominant strategy. Asking for an invoice to prevent tax evasion can therefore be compared to contributing to a public good: government intervention is necessary, as otherwise 'prevention of tax evasion' will remain at a level below the social optimum (e.g. a high level of tax avoidance).

3.4. Giving customers an incentive to ask for an invoice through the Lottery Ticket Reward Policy

Given the findings in the previous section, the question is how to make customers ask for an invoice. In some countries, customers could face sanctions if they did not ask for an invoice. This was the case in Belgium and Italy. In Italy the obligation to issue an invoice was introduced in the 1980s. Orig-

inally, sanctions were imposed both on non-compliant business owners and customers. However, in practice it was problematic to impose sanctions on customers. The sanctions were strongly criticized by the population and the public opinion. The main reason was the high number of sanctions imposed on ignorant customers as a consequence of buyers' mistakes[44]. Moreover, customers had the troublesome duty of storing invoices for a period of time. These factors generated in the population a feeling of resentment against the monitoring authority and not only proved ineffective in fighting tax evasion, but seemed even to produce countervailing effects. As a consequence, in 2003 the Italian government abolished sanctions on customers[45]. Similarly, sanctions on buyers that did not request an invoice were in place in Belgium for a while but they were difficult to impose, were mainly symbolic and have been abolished as well.

An alternative to sanctioning customers is to give them a reward if they ask for a receipt. However, it might be rather costly and lead to heavy administrative burdens to give each customer a cash reward. Furthermore, if the reward is not high enough, customers will not be induced to ask for an invoice. For example, in the 1980s Bolivia tried to encourage people to require VAT receipts by introducing a complementary withholding tax of 10% on all income, which could be offset against the VAT paid as verified by invoices. However, according to Bird it was far from clear that this device boosted tax enforcement significantly, one of the reasons being that the stimulus to collect receipts was weak given the alternative of making a deal with the entrepreneur not to pay the VAT and splitting the difference (Bird, 1992).

Instead, countries can give customers who ask for an invoice a chance to obtain a large reward. This is not only cheaper, but Alm et al. (1992) show in

[44]Newspapers often emphasized cases where sanctioned buyers were children or where the sanction was the consequence of a accidental mistake (see for example *Corriere della Sera*, 18 March 1998).

[45]D.L. n. 269 (2nd October 2003)

a laboratory experiment that rewarding tax compliant behavior with participation in a lottery increases the rate of compliance more than rewarding all compliant individuals. In order to implement this reward policy, the government starts a lottery. Each invoice issued becomes a lottery ticket by way of a serial number that is printed on every invoice. Hence, in order to participate in the lottery, customers have to request for an invoice and keep it until the final draw. The winning numbers are drawn from all serial numbers and the individuals owning the invoices with the winning serial numbers can claim a prize. If the costs of organizing the lottery and of paying out the prizes are smaller than the increase in tax revenue, the government increases its final tax revenue at zero cost. Furthermore, the lottery might have the effect that customers become so used to asking for a receipt that over time prizes may decrease in value or eventually be abolished. Thus it could be a means of strengthening tax morale in a country. This reward policy is also known as the Lottery Ticket Reward Policy (in short: LTRP).

While formal analysis of this topic started only in recent years, the idea of using lotteries and contexts in order to finance public goods is not a novelty. For example according to Karoshi (2008) already during the Chinese Han dynasty (205 – 187 B.C.) the construction of the Great Wall of China has been partially financed through lotteries. The seminal contribution in economic literature is due to Morgan (2000). The author theoretically analyzes the performance of lotteries and raffles compared to voluntary contribution in the private provision of public goods. He sets the conditions under which lotteries outperform voluntary contribution mechanism, finding that the degree of efficiency obtained is an increasing function of the prize size. After Morgan's contribution, several papers have sought to confirm and further investigate his findings through laboratory (Carpenter et al., 2010; Corazzini et al., 2010; Faravelli and Stanca, 2012; Lange et al., 2007; Morgan and Sefton, 2000; Orzen, 2008; Schram and Onderstal, 2009) or field experiments (Landry et al., 2006; Onderstal et al., 2011). A common finding in this

literature is that fixed-prize lotteries or auction mechanisms outperform voluntary contribution mechanism in the private provision of public goods[46]. However the focus of these studies is on which mechanism for awarding the fixed-prize works the best (lotteries vs. auctions; single- vs. multi-prize lotteries; first-pay vs. all-pay auctions; etc.), independently from the capability of the fundraising mechanism to finance itself the value of the prize. Indeed results reported by Landry et al. (2006) and Lange et al. (2007) show that in their environments individual contributions were insufficient to cover the fixed-prized value.

For the purposes of the present paper, the ability of LTRP to self-finance itself is a key issue. The only contribution investigating the performance of lotteries compared to voluntary contribution mechanism on the private provision of public goods under the condition that the public good provision must be self-financing is Duffy and Matros (2012). The authors consider an environment where the public good is provisional instead of exogenously given. That means the public good is created only if the total contribution collected is greater or equal than the lottery prize value, otherwise the public good is not provided and individual contributions are returned. In a laboratory experiment, the authors show that a set of conditions exists for which a fixed-prize lottery incentivizes participants to positively contribute to the public good and that participants' total contribution exceeds the value of the lottery prize.

Therefore, according to theoretical predictions and empirical evidences coming from field and laboratory experiments, it is possible to increase the private provision of public goods by means of self-financing lotteries. Hence, if the

[46] An exception are results reported by Onderstal et al. (2011). In a field experiment comparing different charity fundraising mechanisms, the authors find that voluntary contribution mechanism raises the most money followed by fixed-price private value lottery and fixed-price private value all-pay auction mechanism. The authors conjecture that the prize offered in the lottery and auction treatments may have crowded out intrinsic motivation to contribute to the charity among the participants.

parallel between asking for an invoice and a public goods dilemma is correct, the lottery mechanism underlying the LTRP could be exploited in order to enforce invoices emission. In the next section I provide a model explaining the mechanism on which the LTRP is based.

However, LTRP is not just a theoretical approach to combating VAT and RST evasion, it has actually been implemented in several countries. Taiwan implemented such a reward policy in 1951 which is called the Uniform-Invoice Prize Winning Lottery. After the introduction of the uniform invoice system in Taiwan, it turned out that firms tended to underreport sales by not issuing an invoice at the time of sale. The tax authorities tried to induce customers to ask for invoices with every purchase. Most importantly, this kind of behavior was being induced by the uniform-invoice lottery giving customers the chance to win a large amount of money by obtaining an invoice at the time of purchase (Lin, 1992). Every one of the roughly 11.5 billion receipts issued annually by Taiwanese shops comes with a unique lottery number, which enters a bi-monthly prize draw awarding prizes of up to $ 342,000[47]. Customers can check on line whether they have won a prize[48]. This policy is still in place, according to Giebe and Schweinzer (2013) because it proved so successful. Some other countries that have applied the LTRP are the Philippines, Malaysia, Chile, Puerto Rico and Brazil. According to Giebe and Schweinzer (2013) these schemes have been highly successful in their intended purpose of reducing tax evasion.

Recently, LTRP has been applied in some European countries as well. Portugal introduced a peculiar version of LTRP in 2014, where the prizes paid out to lottery winners are luxury cars. Also Slovakia adopted the policy in 2013 and the first data shows that more than 450,000 people took part to the lottery registering more than 60 millions receipt and comparative statics

[47]Giebe and Schweinzer, 2013 and `http://en.wikipedia.org/wiki/Uniform_Invoice_lottery`.

[48]`http://www.etax.nat.gov.tw/etwmain/front/ETW183W6?site=en`

shows a huge increase of VAT collected compared to the previous year. Furthermore, since the Slovakian version of the LTRP allows private citizens to verify if the business owners correctly registered the transaction that originated the invoices, notifications to the tax authority of tax evading business is 25 times higher compared to the pre-LTRP introduction period[49].

Despite these practical experiences, until recent years there was nothing more than descriptive statistics and anecdotal evidence for these positive results. No systematic analysis was conducted on the impact of LTRP implementation. One of the reasons for this might be the technical difficulty in isolating the causal effect of a policy introduction. If a policy is adopted at a state level, it would be complicated to find a credible comparison. A suitable comparison could be another country that didn't implement the policy but that is otherwise similar to the country that did introduce it, but it is difficult to find comparable countries. Cross-country comparison results are often considered to be unreliable.

However, since 1998 a peculiar implementation of the LTRP in China makes it possible to isolate the causal effect of the policy. At that time, one of the turnover taxes levied in China was the so called business tax (BT), a turnover tax levied mainly on specific services. This tax was generally collected by local tax authorities. In order to reduce the negative effects of widespread BT evasion, the Chinese government started printing a lottery number on receipts registering business transactions. The invoice for restaurant or entertainment expenditures is at the same time a lottery scratch card. The idea is that customers will be incentivized to ask for an invoice and thus oblige the service provider to pay BT. Each lottery pays out a prize after some period of time. Once the receipt is issued, the seller cannot evade BT on that transaction. Thus, the buyer has a direct incentive to ask for the receipt and this indirectly obliges the seller to reveal information to the

[49] *In Slovakia, Real Lottery Prizes go to Tax Men*, New York Times, April 19th 2014.

tax authorities. The peculiarity of the Chinese experience is the particular form in which the LTRP was implemented. The Chinese State Commission for Restructuring the Economic System[50], a Chinese governmental agency, decided to introduce the LTRP only in some experimental districts in the period 1998 - 2003 in order to test its effects. At first, only some service industries, such as food service businesses, issued lottery tickets. As of 2002, the LTRP was applied to other service industries as well. Furthermore, the trial area was expanded to involve a growing number of districts. Because of this isolated implementation of the LTRP, it is possible to compare relatively similar districts with and without the LTRP. Therefore, the Chinese experience is a (quasi-) natural experiment.

There has only been one study conducted by Wan (2010) that investigates the effects of this policy in China. Wan estimated that the lottery reward policy increased revenues from BT by 17% in the experimental districts. He estimated that the ratio between lottery prizes paid by the government and increased tax revenue ranged between 1:30 and 1:40. This success induced the Chinese government to extend the LTRP area progressively from the initial trial area to the whole country[51].

[50]See Note of Mainland China Government by State Commission for Restructuring the Economic System, 1989.

[51]However, a word of caution on the implementation of the lottery policy in China is necessary. Some scattered data collected in China during the experimental period show that at the time of the lottery draft the Chinese government paid out only a relatively small fraction of the announced prizes. For example, while the Beijing Local Tax Bureau announced that prizes would amount to thirteen million Yuan in 2002 (see Beijing Local Tax Bureau announcement on July 17th 2002) ex-post payments are on average less than 17% of the prizes previously announced. Such inconsistent behavior maximizes revenue in one period but, needless to say, would kill any possibility of collecting revenues in succeeding periods as soon as customers find out that prizes are not actually paid. Given the lack of comprehensive data on this issue and the relatively short experimental period, future research should test whether the success of the policy in the first years decreased over time. In this chapter, I will focus on the explanations for the success of the lottery policy in the initial stages, in which consumers expected prizes to match those previously announced.

Understanding the determinants of the successful results of the LTRP is not merely a theoretical exercise but a key element in effectively replicating the policy in different contexts. After having decided to implement LTRP, a government has to commit to pay a lottery prize to the winner of the lottery. If the ex post increase in tax revenue is smaller than the prize, the government incurs a loss. A theoretical model that captures and explains the key factors involved in the LTRP mechanism would provide an indicator of the likelihood of success in a specific socio-economic and institutional environment. That would limit the probability of unsuccessful implementation of the policy and possibly prevent monetary losses for the government.
In the next section I present a model based on non-expected utility that explains the success of LTRP and that will help policy analysts considering LTRP implementation in predicting the policy outcome.

3.5. The Model

Consider a public goods situation where it is not possible or feasible to increase the level of private contribution by increasing sanctions. Define parameters as:

N : number of players.
t : 1,..., T : number of periods in which it is possible to contribute to the public good.
y_i : initial endowment player i.
x_i : expected payoff player i.
$a_{i,t}$: per period contribution player i.
a^* : per period required level of contribution to get a lottery ticket (exogenously settled).
m: marginal per capita return to the public good.

Under a voluntary contribution mechanism to the public good with no lottery the expected individual payoff for each period is:

$$x_i = y_i - a_i + m \sum_{j=1}^{N} a_j \tag{4}$$

In order to replicate a public goods game situation set the parameters in such a way that it holds:

$$\begin{cases} m > (1/N) \\ m < 1 \end{cases} \tag{5}$$

Participants to the public goods game maximize individual payoffs with respect to the chosen contribution level:

$$\frac{\partial x_i}{\partial a_i} : -1 + m < 0 \tag{6}$$

Hence, while it is a dominant strategy for individuals to completely free-ride, it would be Pareto-efficient if everyone contributes the full endowment to the public good. Indeed, theoretical predictions indicate that the contribution rate would converge towards a suboptimal equilibrium level of total contribution $\hat{A}$ (equal or close to 0 if it is assumed that a small fraction τ of players always adopt strictly altruistic behavior):

$$\begin{cases} \hat{A} = \sum_{i=1}^{N} a_i = \tau N a^* \\ \tau \simeq 0^+ \end{cases} \tag{7}$$

Now assume that a Central Authority interested in increasing the amount of contributions to the public good collected introduces a lottery linked to the public good, with the prize $\delta_{t=z}$, z = 0, T, 2T,..., cyclically announced at time t and assigned after period T. Each lottery ticket has a probability of being drawn of 1/(N*T), while each subject has the possibility of acquiring a lottery ticket in each period, providing a contribution to the public good $a_i \geq a^*$. Therefore, the individual probability p_i of winning the lottery prize

depends on the individual player's choices of contribution:

$$p_i = (NT)^{-1} \sum_{t=1}^{T} c_{i,t} \tag{8}$$

where $c_{i,t} = 1$ if $(a_{i,t} \geq a^*)$ and 0 otherwise.

The individual per period payoff when the lottery policy is implemented becomes:

$$x_i^R = y_i - a_i + m \sum_{j=1}^{N} a_j + (1/T) p_i \delta_z \tag{9}$$

where δ_z is equal or smaller to the estimated quantity $\hat{\delta}_0$ announced at the initial period for t=0 and paid after T periods; while for t>0 δ_z is equal to or smaller than the total public good contributions collected in the previous T periods after subtraction of the sum of per period voluntary contributions level $\hat{A}$ that is collected when no lottery policy is in place ($\hat{A}$ is assumed to be constant).

Moreover to complete the feasibility constraint the Central Authority takes into account that the lottery prize will not be paid out with probability (1-p^*), where p^* is the fraction of the total number of tickets emitted for each lottery that has been acquired by contributors. Hence:

$$\hat{\delta}_{z=0} = [\sum_{t=1}^{T} \sum_{i=1}^{N} a_{i,t} - T\hat{A}]p^* \tag{10}$$

$$\delta_{z>0} = [\delta_{z-T} - T\hat{A}]p^* \tag{11}$$

$$p^* = (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} c_{i,t} \tag{12}$$

Without loss of generality, assume that in each period agents face a single

binary decision either to positively contribute to the public good or free-ride, hence a_i=0 or a_i=a^*=1. As discussed in the previous section, in the specific context of sales tax evasion a_i could be interpreted as the opportunity cost of a lost price discount combined with the moral costs of requesting a receipt. Moreover for simplicity consider the case where t=T=1 and z=0. Per period p_i becomes:

$$\frac{1}{N} \tag{13}$$

when a_i=1 and 0 otherwise.

Now consider the individual choice over the binary alternative to either contribute or free-ride. Individuals per period payoff given no contribution becomes:

$$x_i = y_i + m \sum_{j \neq i, j=1}^{N} \tag{14}$$

Instead the payoff associated with a contribution to the public good that implies the possibility to win the lottery prize is:

$$x_i^R = y_i - 1 + m \sum_{j=1}^{N} + U(\hat{\delta}, p) \tag{15}$$

To further simplify the analisys and without loss of generality, assume that N is large enough to make the individual contribution is negligible with respect to the quantity of public good provided:

$$\sum_{i=1}^{N} \simeq \sum_{j \neq i, j=1}^{N} \tag{16}$$

The disequation reduces then to compare the value of the prospect $\hat{\delta}$ with that of the required contribution a_i. Hence individual contribute to the public

good iff:

$$U(\hat{\delta}, p) \geq 1 \tag{17}$$

Now introduce heterogeneity in population types. Assume that a fraction (1-ψ), $\psi \in [0,1]$, of the population behaves as an expected Von Neumann-Morgenstern (VNM onward) utility maximizer (Von Neumann and Morgenstern, 1944). Hence individuals evaluate the probabilistic prospect $\hat{\delta}$ through maximization of expected utility. Given the probability to win the lottery prize specified in (9), individuals contribute iff:

$$U(\hat{\delta}, p) = \frac{U(\hat{\delta})}{N} \geq 1 \tag{18}$$

where $\frac{\partial U}{\partial \hat{\delta}} > 0$ and $\frac{\partial^2 U}{\partial \hat{\delta}} \leq 0$.

Proposition 4. *For any VNM expected utility maximizer agent having a utility functional form that does not imply risk seeking behavior the individual optimal strategy of contribution is a_i=0, that is never enforcing invoices emission irrispectively of the implementation of LTRP.*

Proof. Consider the feasibility constraint in setting the prize $\hat{\delta}$ in (6) and the condition for contribution in (14). Furthermore, consider the extreme case of a risk-neutral agent interested in maximizing wealth, and the best-case scenario in which all members of the population contribute to the public good. The condition for individual contribution becomes:

$$\frac{1}{N}(\hat{\delta} = \sum_{a=1}^{N} a_i \in [0, N]) - \hat{A} < 1 \tag{19}$$

Based upon the assumption that utility is marginally constant or decreasing in wealth and risk-seeking preferences are ruled out, the case considered represents the most attractive possibility for a VNM expected utility maximizer agent to accept the gamble opportunity. Hence, given that any other

possible combination of risk-preferences and utility functional form results in a decreased value of the left side of equation (15), it is possible to conclude that free-riding remains the dominant strategy for VNM-type agents, independent of the implementation of the LTRP. □

Now assume that the remaining fraction ψ of the population evaluates the prospect $\hat{\delta}$ through Cumulative Prospect Theory (CPT), which is a model describing decisions under risk, proposed in their path-breaking articles by Tversky and Kahneman (1992). The theory was introduced in order to capture some behavioral regularities in individual decision-making, such as risk-seeking, loss aversion and the overweight (underweight) of unlikely (average) events, which could not be explained by Expected Utility Theory. In particular, CPT modifies Expected Utility Theory by replacing final wealth with payoffs relative to the status quo, replacing the utility function with a value function that depends on relative payoff, and replacing cumulative probabilities with weighted cumulative probabilities.

For the purpose of this chapter, the interesting aspect of CPT is the attention paid to behavioral regularities (or anomalies, from the perspective of VNM) such as nonlinear preferences and risk-seeking behavior in betting and lotteries[52]. In fact, it is well-known that Expected Utility Theory cannot explain why individuals buy insurances and at the same time like gambling (Camerer et al., 2004). For example, according to the EUT, a rational agent who prefers \$800 as a certainty over the prospect of \$2000 at a probability of 50% will also decline the prospect of \$1,000,000 at a probability of 0.1%, since the two probabilistic prospects have the same expected value of \$1,000.

[52] Additionally summarized by Camerer and Loewenstein (2004b, p.22): "Expected Utility hypothesis is like Newtonian mechanics [...]. Linear Probability weighting in Expected Utility works reasonably well except when outcome probabilities are very low or high. But low-probability events are important in the economy, in the form of "gambles" with positive skewness (lottery tickets, and also risky business ventures in biotech and pharmaceuticals), and catastrophic events that require large insurance industries. [...] People are typically averse to risky spreading of possible money gains."

However, contrary to EUT predictions, the empirical evidence shows that a consistent percentage of the population systematically prefers the certainty of \$800 to a 50% chance of obtaining \$2,000 (showing risk-aversion) but at the same time would also prefer the prospect of winning \$1,000,000 at a 0.1% probability or even at 0.01% probability in preference to the certainty of \$800 (hence even showing risk-seeking in the last case, since the expected value of the probabilistic prospect in the latter case is smaller than \$800). CPT explains by way of a formal theory the empirical evidence that individuals systematically do not maximize expected utility when facing probabilistic prospects with certain characteristics: they instead overweight the likelihood of extreme events and remain relatively unaffected by changes close to the average of the probability range.
Specifically, CPT implies that individuals non-linearly weigh the probability of gaining the lottery prize and evaluate the lottery outcome by means of a value function. In the discussion that follows, I adopt the same value and weighting functional forms proposed by Tversky and Kahneman (1992)[53].

Proposition 5. *When LTRP is established, it is possible to find a set of conditions under which, for a CPT-type agent, contributing to the public good and so enforcing invoices emission becomes the individual dominant strategy.*

Proof. Consider CPT value and weighting functions as presented in Tversky and Kahneman (1992). Substituting and plugging the terms defined above in CPT model it is possible to derive individual conditions for contribution:

[53] A correct characterization of the most appropriate probability weighting function and value functional form and a calibration of the functions' parameters is beyond the scope of the present work. In this chapter, I adopt a polynomial value function and the parameters value reported in the original Tversky and Kahneman article. However, it should be underlined that the estimation of the correct value function and the calibration of the parameters remains an open issue. Nevertheless, note that it is possible to show that the qualitative results obtained by assuming a polynomial utility function hold for any other continue and quasi-concave functional form.

$$U(\hat{\delta}, p) = \frac{(\frac{1}{N})^{\sigma}}{((\frac{1}{N})^{\sigma} + (1 - (\frac{1}{N})^{\sigma}))^{\frac{1}{\sigma}}} \hat{\delta}^{\rho} \geq 1 \tag{20}$$

where $\sigma \in [0,\ 1]$ and $\rho \in [0,\ 1]$ describe respectively the curvature of the weighting function and the degree of risk aversion.

Since the fraction (1-ψ) of the population constitutes of VNM-type agents will not contribute to the public good for any feasible amount of $\hat{\delta}$, the feasibility constraint for the Central Authority setting the prize becomes:

$$\psi\hat{\delta} \leq \sum_{i=1}^{N} a_i - \hat{A} \tag{21}$$

because with probability (1-ψ) the lottery prize remains with the Central Authority. Substituting $\hat{A}$ according to (7), rearranging (20), and solving for $\hat{\delta}$ restricting the attention to the case of equality the result is:

$$\hat{\delta} = N - \frac{\tau N}{\psi} \tag{22}$$

Plugging (22) in (20) results in a non-linear equation characterized by the parameters N, τ, ψ, ρ and σ. Hence it is possible to derive the condition under which it becomes a dominant strategy for individuals to contribute to the public good:

$$U(\delta) = \frac{(\frac{1}{N})^{\sigma}}{((\frac{1}{N})^{\sigma} + (1 - \frac{1}{N})^{\sigma})^{\frac{1}{\sigma}}} (N - \frac{\tau N}{\psi})^{\rho} - 1 \geq 0 \tag{23}$$

Given the population size and the value of the parameters ρ and σ, disequation (23) is greater than 0 when $\frac{\tau}{\psi}$ is sufficiently smaller than 1. This is equivalent to say that, when the number of agents requesting an invoice even absent the LTRP is relatively small compared to the number of CPT-type in the population, it is possible to offer a prize $\hat{\delta}$ such that requesting for an invoice becomes a dominant strategy for all the CPT-type agents. □

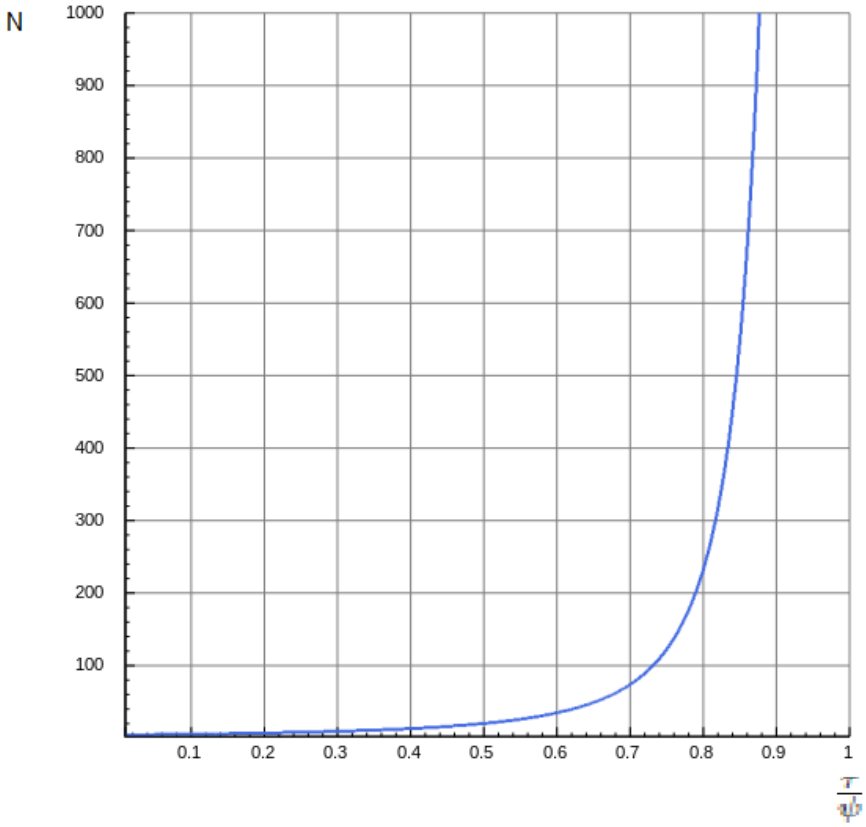


Figure 3: Value of the ratio between proportions of agents enforcing invoices emission absent LTRP and CPT-type in the population allowing for setting a LTRP prize such that a CPT-type agent is indifferent whether to enforce invoices emission or not and the LTRP is self-financed.

How small does the fraction of the population enforcing invoices emission even absent the LTRP compared to the fraction of CPT-type agents has to be? In Figure 3, I graph the value of the ratio $\frac{\tau}{\psi}$ that allows setting a LTRP prize $\hat{\delta}$ such that a CPT-type agent is indifferent whether or not asking for an invoice, as a function of the population size N[54]. Therefore, given the population size of the situation the policy analyst is considering, $\frac{\tau}{\psi}$

[54]I assumed ρ=0.88 and σ=0.61, as estimated by Tversky and Kahneman (1992).

Table 4: Upper bound $\frac{\tau}{\phi}$ allowing self-financed LTRP reward that makes agents indifferent between enforcing invoices emission or not for given population levels (σ and ρ values taken from Tversky and Kahneman, 1992).

Population size	100	1,000	10,000	100,000	1,000,000
Ratio $\frac{\tau}{\phi}$	0.731	0.877	0.940	0.970	0.986

reported in Figure 3 represents the upper bound for the ratio $\frac{\tau}{\psi}$ that makes implementing the LTRP policy possible without earning negative expected profits. For any value of the ratio $\frac{\tau}{\psi}$ higher than this upper bound, the cost of the prize paid out by the LTRP would exceed the increase in revenue collected. Table 4 reports the exact value of this upper bound for some population sizes.

3.6. Discussion of the Results

From what I showed above, it follows that, for any feasible prize amount offered by the government, the individual dominant strategy for a Von Neumann-Morgenstern Expected Utility maximize agent with any non-negative degree of risk aversion, remains not to request an invoice. Individuals evaluating the probabilities of winning the lottery prize multiplied by the prize amount will always find that the expected gain deriving from the lottery is smaller than the cost of asking for the receipt. Hence, if individuals behave as a Von Neumann-Morgenstern utility maximizer, the LTRP would result in a failure unless it is unrealistically assumed that individuals are risk-lovers.

Therefore, in light of the evidence of the success of the LTRP discussed in previous sections, it seems that Expected Utility Theory is not the appropriate theoretical background to analyze or explain individual decision making in the context of the LTRP. The reason is that the LTRP introduces

a probabilistic situation in which individuals choose over extreme elements: an extremely low probability of winning a substantial prize. For this kind of situation the linearity in probability weighting implied by Expected Utility Theory seems unable to capture the underlying decision-making process.
CPT instead represents a more suitable theoretical background to analyze the LTRP. Implementing the theoretical framework of CPT allows for making ex-ante predictions on the successful implementation of the LTRP. In particular, once information about risk preferences and size of the population of interest has been collected, a policymaker could determine if the lottery prize associated to LTRP would be sufficiently large to persuade CPT-type individuals to enforce invoices emission.
To proceed with this calculation it is necessary to acquire information on the gambling and risk preferences of the population. In practice, it is necessary to generate a quantitative description of the agents' average behavior when facing decisions under risk. To be technically precise, it is necessary to calibrate the parameter values of the model adopted in describing individuals' behavior under risk and uncertainty. The successful implementation of the LTRP in China does not guarantee that the same policy would achieve equal results in a different environment, since it is well known that individuals' risk-preferences greatly vary across societies. Given that many observable (such as income per capita or average saving rates) and unobservable (such as culture and social norms) factors are correlated with the taste for gambling of a population, establishing the possibility of a successful implementation of the LTRP in a specific environment requires a careful empirical investigation of the characteristics of a population.
The verification whether a country with a higher level of income per capita and different ethical norms than China shares a taste for gambling sufficiently developed to implement the LTRP is an empirical issue. In order to clarify how this estimation of population's gambling behavior works in practice, consider the situation in which a government wishes to apply LTRP. Before

announcing the lottery prize, the government will want to check if the policy described in an abstract context will work in this specific country. As a first step, a quantitative characterization of the risk preferences of the population has to be estimated. Statistical procedures and econometric techniques may fulfill this task (see Andersen et al., 2008 for a detailed discussion of this point). While a detailed discussion of these methodologies lies outside the scope of the present chapter, it is useful to provide some examples. Survey results and field data relative to lottery tickets sold could be used to estimate the average part of income spent on lotteries and on gambling (Harrison et al., 2007). Alternatively, it may be possible to directly elicit the risk-taking preferences of representative random samples of individuals through interviews or small incentivized acts of gambling. A detailed discussion of this last procedure, commonly used in experimental social sciences, is reported in Holt and Laury (2002). Once a quantitative characterization of the population's risk preferences is obtained, it would be sufficient to incorporate those values into the model presented above. Then it can be established whether, given the estimated risk preferences, the population of the specific country is large enough to attempt a successful implementation of the LTRP.

3.7. Possible Counter-arguments

The empirical evidence discussed above and the model presented in the previous section suggest that the LTRP could be an effective tool for policymakers to achieve socially efficient outcomes. Nevertheless, a possible counter-argument is that the policy requires a government at time zero to commit to paying an ex-ante announced high monetary premium. However, the effective increase in tax revenue only occurs later. The prize amount initially offered could be seen as an investment that can only partially guarantee future returns as it is made under conditions of uncertainty. While the policy is founded on a theoretical argument supported by experimental and empirical evidence, the practical implementation and design of such a reward mechanism in real-life environments could be extremely complex and subject to

failure.

Moreover, in some cultures there might be a moral aversion to lotteries, which will make it politically difficult to implement the policy. When the Belgian Minister of Finance only hinted at a lottery system for restaurant and bar invoices in December 2009, it was immediately criticized by a Member of Parliament. She seemed to fear that it might lead to a gambling addiction[55]. Furthermore, the mechanism rests on the assumption that people's taste for gambling will not decrease over time. It should be tested if individuals' willingness to ask for invoices boosted by the excitement about the new gambling opportunity in the periods immediately after the reward policy has been implemented are followed by a progressive decline in interest (and in the request for invoices) over time. Sustainability of the lottery ticket policy in the long-run depends crucially on this factor[56]. For example Bird (1992) is skeptical about what he calls 'tax gimmicks' as the LTRP. In his view the real secret of success lies not in such gimmicks but in the more mundane task of establishing a more credible and effective tax administration. Bird acknowledges that if tax administration is improved, then 'gimmicks' intended primarily to increase the flow of information to the administration may provide some extra benefit, but in his view these cannot take the place of improved administrative effort. I agree with Bird that improving the tax administration is extremely important to improve compliance. However, for countries that do not have the means and knowledge for bringing their tax administration up to the highest standard, policies such as the LTRP might be of help.

Also, when developing an LTRP, mechanisms must be introduced to reduce fraud with invoices, such as falsified invoices. In Taiwan new systems of e-invoices which are being proposed include the special function of auto-

[55]Réponse du vice-premier ministre et ministre des Finances et des Réformes institutionnelles du 08 mars 2010, à la question n° 270 de madame la députée Valérie Déom du 07 janvier 2010, DO 2009201013743, QRVA 52 97, p. 82-83.

[56]As noted above, data on the results of the natural experiment occurring in China are available only for a relative short period of time.

matically checking whether the invoice number matches the Uniform-Invoice Prize Winning Numbers announced by the Ministry of Finance (Chang et al., 2012). Such systems will also help to reduce falsification of VAT receipts. Another problem with the Taiwanese system was the fact that as the lottery numbers come per invoice and not per amount spent, there is an incentive for customers to pay for every single item separately in order to get more receipts (Giebe and Schweinzer, 2013). A possible solution for this specific problem would be paying a lottery prize that is proportional to the invoice value. This solution would drop customers' incentives to pay for each item separately, since the increase in probability of winning the lottery due to the fact that the buyer collected multiple invoices is offset by the diminished value of the lottery prize.

Furthermore, it has been suggested that targeted rewards may be more effective than scattergun rewards. Giving the chance to win lottery prizes to all customers may not seem as effective as rewards to specific customers, such as customers who report painters who offer them a discount for cash with no invoice. While it is true that this mechanism could potentially increase the lottery efficiency compared to LTRP, nevertheless, the practical implementation may also bring additional problems. A system that rewards only customers who actively report irregular transactions implies that the individual reporting the illegal action has to reveal personal data. This could potentially restrain customers who want to remain anonymous when reporting illegal actions of sellers. For example, in Italy customers can report to the Guardia di Finanza, the official monitoring authority, irregularities in the issuance of invoices (in 2012 there have been more than 600.000 notifications). On the basis of this information, the authority may decide to impose an audit on the targeted business. While before 2012 notifications were strictly anonymous, starting from April 2012 the Italian government required personal data from the customer reporting the irregularity. This decision of the Italian legislator provoked criticism since customers reporting

irregularities could be identified and have often been subject to material and moral retaliations. It is still difficult to empirically assess the effects of the government policy. However, anecdotical evidence suggests that because of it many customers reporting irregularities in the issuance of invoices switch from the official Guardia di Finanza signaling system to an unofficial website (www.evasori.info) created by a private citizen in order to report tax evasion anonymously.

Finally, a special word of caution should be spent on the crowding-out effect of voluntary requests for invoices. In some countries, a consistent percentage of the population considers it to be an individual duty to enforce the issuance of invoices, even without specific laws or monetary incentives. Unfortunately, those customers who regularly request invoices may not carry on doing so after LTRP is introduced. There is a growing body of literature both in psychology and economics focusing on the direct and indirect detrimental effects of monetary incentives (see, among others, Frey and Jegen, 2001; Le Grand, 1997; Benabou and Tirole, 2003). Those studies suggest that monetary incentives directly crowd out individuals' willingness to behave pro-socially. Furthermore, these studies suggest that these incentives indirectly affect the proper functioning of a norm enforcing mechanism, increasing inefficiency.

Investigating this issue, Fuster and Meier (2010) set up a laboratory experiment in order to verify the presence of the negative indirect effect of monetary incentives. In each period, participants could allocate a fraction of their private endowment to a public account. Money on the public account generated interests that were distributed at the end of each period. However, interests and capital on the public account were equally shared among all participants, independent of their individual contribution. This scenario mimics a public goods situation: while it would be socially efficient for participants to allocate the full private endowment to the public account, the individual dominant strategy consists in free-riding on others' contribution. As previously discussed, it is well known that without any external intervention,

the level of resources allocated to the public account remains sub-optimal. However, despite the theoretical prediction of zero contribution, it has been shown that a proportion of participants always adopt the socially efficient strategy, irrespective of what the other players are doing. The objective of Fuster and Meier's experiment is to verify the effect of a monetary reward on the behavior of these altruistic participants. When monetary incentives for adopting socially efficient behavior are introduced, altruistic agents did not always carry on behaving consistently. Instead, while a number of free-riders started behaving pro-socially because of the incentives, some of the altruistic agents stopped allocating resources to the public account. In the end, the combination of these effects leaves the net amount collected on the public account unchanged in the situation with or without the private incentives scheme. The possible explanation for this counter-intuitive and inefficient result suggested by the authors is the destruction of intrinsic motivation by extrinsic incentives and the framing effect of shifting from a social to a monetary context.

Fuster and Meier's results are important for the LTRP. These suggest that LTRP could be effective and self-sustaining, leading to a stable, efficient, equilibrium, only if a series of fundamental accessory conditions is present. Specifically, it seems that the possible crowding out effect of monetary incentives on norm enforcement would not be a problem in the case of widespread and inefficient socially accepted behavior, such as tax evasion and not asking for an invoice. In situations with established inefficient social norms little altruistic enforcing of the issuance of invoices is to be expected without government intervention. Thus, a well specified system of incentives could achieve a higher contribution level without leading to negative indirect effects.

3.8. Positive Long-term Effects

Despite the concerns emerging from possible side-effects, there are also positive externalities connected to the lottery policy. First of all, imagine the

LTRP is introduced in a society where tax evasion, in the form of not issuing invoices, is widespread and that this behavior is socially accepted or tolerated. If the LTRP is adopted, it is reasonable to assume that some consumers will now react to private incentives and will start enforcing the issuance of invoices even from suppliers that were used to systematically evade tax. The negative aspects of the social costs of asking for a receipt are outweighed by the chance of winning a prize.

Through the historical records of VAT or RST reported by companies, the tax authorities can identify those businesses that have an abnormal peak in the period in which the lottery policy is implemented. For example, it would be straightforward to implement an algorithm that, after controlling for seasonality and business cycles, automatically identifies the suppliers reporting a statistically significant increase in supplies and tax. Hence, it would become possible to separate such businesses from those that present continuous payments of VAT or RST. This signal could be used as an indicator to direct monitoring resources towards businesses that report discontinuous trends. Thus, the LTRP could be of help in focusing auditing efforts. Businesses that were used to evade taxes might even anticipate the increased probability of an audit and will review their behavior and increase their VAT or RST payments permanently. As discussed before, it is possible that LTRP will turn out to be unsustainable because the increased payments of VAT and RST are not sufficient to pay the promised prize. If this happens the government will have to incur a momentary loss. However, the benefits of higher contribution levels deriving from more efficient screening and auditing and a more effective sanctioning system will also produce a revenue increase in subsequent periods when the lottery reward option has been abolished.

Moreover, the LTRP may not only be effective in combating VAT and RST evasion, but also in tackling the evasion of taxation of business profits. As invoices give an indication of retail sales, these can be used to establish whether the reported taxable profit is consistent with such retail sales.

Finally, an additional long-term possible benefit deriving from LTRP introduction is the so called equilibria shift in a no pain no gain situation. Following Parisi (2000), we could interpret the apparently irrational presence of Pareto-inefficient social norms (consumers accepting the evasion of tax by their suppliers) as a point of local optimum that requires an initial loss of utility to shift toward the global optimum. To clarify this point, consider as an example the release of more efficient software. This new software is not essential to perform fundamental operations, but individuals using the old software are slower in performing certain minor tasks. Hence, while individuals are not obliged to use the new software, sticking to the old one they experience small disutilities that could be potentially eliminated, resulting in a Pareto improvement (the "gain"). However, utilizing the new software requires a training period during which it is not possible to conduct work activities and an initial effort to learn the new code (the "pain"). If individuals are not sufficiently forward looking (technically, are characterized by a high time discount factor) or don't have information about the benefits of adopting the new software (are rationally bounded), they will refuse to incur the once-and-for-all switching cost to the new software and lose the chance of a permanent improvement.

Similarly, a society as a whole could experience a permanent Pareto improvement if tax revenue increases and the state can provide better services. The change of a social norm fostering tax evasion would be perceived only as a cost in the short run, since less cash would circulate in the economy and less competitive businesses would be likely to fail. Permanent benefits from a change in the status quo will be experienced only in the medium and long run, after the new equilibrium is reached. For example, if the increase in tax revenue is used to finance new infrastructure, only after the project is completed will individuals experience an increase in utility. The introduction of a lottery reward could work as a sort of compensation for the initial "pain" that customers have to experience. Once the new, Pareto superior equilib-

rium is reached, individuals will perceive the enforcement of the issuance of invoices as the welfare-maximizing strategy, even if the LTRP is suspended. Moreover, the external shock could lead to more consumers adopting socially efficient behavior (asking for invoices) and thus initiate a process of changing the norm. The mechanism of social norm creation is often characterized by the so called "snowball effect": an initial group of individuals adopting socially efficient behavior because the external incentives might prompt the rest of the population to ask for invoices as well (Aviram, 2004). Even if, after the first prize is assigned, the government cannot repeat the lottery, it is still possible that consumers will have already reached the new, Pareto-efficient equilibrium and will, therefore, continue to ask for invoices. Asking for an invoice will thus have become the social norm. While it is possible that the initial investment and incentives mechanism will last for only for a limited amount of time, the positive externalities may continue to spread into the future.

3.9. Conclusions of chapter 3

The implementation of the LTRP in China increased RST revenue by giving customers an incentive to request invoices, thus reducing RST evasion by businesses. In this paper I have tried to explain this result and to provide for a model which might help governments in deciding whether or not to implement such a policy to combat RST evasion. Risk preferences, social norms and population size have been discussed as important factors.

A major concern is the level at which lottery prizes must be set. A well specified reward option must elicit a taste for gambling by consumers and induce them to ask for an invoice even though this is not an efficient strategy for a rational utility maximize individual. Given the peculiar situation introduced by the LTRP (low probability of a possible high gain), in order to describe a situation in which agents have to make a decision under risk a generalized theoretical framework based on Cumulative Prospect Theory has been proposed. This general theoretical framework allows for the testing

of the applicability of the LTRP in specific contexts. A key element from a practical point of view would be the correct estimation of risk-preferences of the specific population. I underlined the importance of this empirical task in order to successfully implement the LTRP, since it is well known that risk-preferences vary across populations and depend on individual wealth and other factors. Moreover, I have discussed the possible positive and negative side-effects. In order to limit the risk of crowding out virtuous behavior, I suggest that the lottery only be introduced in countries with high levels of VAT and RST evasion by businesses and a social norm of consumers not asking for invoices or only in sectors with relatively high rates of tax evasion, in countries which have an otherwise compliant norm. For example, where the LTRP might be effective on a more general scale in Italy, it might be best for the Netherlands to limit it to certain sectors, such as those involving decorators and the carrying out of other odd-jobs for private individuals.
Regarding the positive long-term side-effects, I have pointed out how, in some settings, the LTRP could help in deciding which businesses should be audited and that it could result in asking for invoices becoming the social norm, even if the policy is implemented for a limited time only. The side effect of slightly increased waiting times at Milanese bakeries because every customer demands a receipt and less juicy conversations during Dutch birthday parties about decorators, will be outweighed by such benefits.

4. Social Influence on Third-Party Punishment: an Experiment[57]

Like chapter 3, the present chapter suggests an original contribution to the field of behavioral public policymaking. I investigate the possibility to exploit social influence effects in order to increase bystander altruistic intervention, or using the language of economics, costly third-party punishment. While both social influence and third-party punishment have been extensively investigated by scholars in the social sciences, this is the first study in either law or economics that focus on the their interconnection.

I start proposing a model of social influence and deriving theoretical predictions that diverge from results obtained by neoclassical models of decision-making. I then test my model predictions empirically. In order to isolate the causal effect of social influence on third-party punishment from confounding factors, I rely on the controlled environment of a laboratory experiment combined with the methodology of experimental economics.

Results of my experiments show that social influence is a major determinant of third-party punishment. Moreover, they allow to identify the individual characteristics that make an agent more sensitive to social influence. Finally, I show that some subject respond to normative social influence (the "need to be liked" by peers), but their choices are not affected by informational social influence ("the need to be right").

My findings are relevant for policymakers and decisionmakers that consider the possibility to implement policies based on third-party interventions. I argue that costless and easy to implement social influence approaches would increase the policies effectiveness and achieve welfare-improving results.

[57]I am grateful to the Alfred P. Sloan foundation for financial support. I am deeply indebited to Emanuela Carbonara and Marco Casari for helpful comments and discussions: this chapter would have not be written without their help. I also thank Maria Bigoni, Andrea Geraci, Riccardo Ghidoni, Francesco Parisi, Louis Visscher, Roberto Weber and seminar participants at Erasmus University Rotterdam for helpful suggestions and Stefano Rizzo for valuable research assistance. The usual disclaimer applies.

4.1. Introduction

Societies often rely on punishment for preventing and eventually responding to rule violations. Punishment is a costly activity that inflicts negative consequences upon wrongdoers and that is carried out both by formal centralized institutions and by the decentralized actions of peers. When the punishment activity is inflicted directly by agents that are bearing the costs of the rule violation, we talk about second-party punishment. However, in groups composed of a large number of agents interactions are often non-repeated and the punishment activity is typically carried on by a third-party not directly affected by the consequences of the rule violation. In these cases we refer to third-party punishment.

While scholars' attention has traditionally focused on second-party and centralized third-party punishment, in recent years a growing body of contributions analyzes empirically the role of decentralized third-parties in punishing rules and norms violators (Fehr and Gächter, 2002; Fehr et al., 2003; Fowler, 2005). However, despite there has been substantial progress in identifying the determinants of decentralized third-party punishment (Bernhard et al., 2006; Lewisch et al., 2011; Lieberman and Linke, 2007; Coffman, 2011), there is only a partial understanding of what are its major determinants yet (Fehr and Fischbacher, 2004a).

In this chapter I focus on decentralized punishment, examining the effects of social influence on the punishment behavior of bystanders not directly affected from the action of the wrongdoers[58]. By social influence I refer to the effect of the endogenous interactions between a third-party's preferences for punishment and the preferences for punishment expressed by other bystanders (Manski, 2000). Focusing on endogenous interactions means that I investigate the influence that the punishment choices of other third-parties have on the decision of a bystander to engage in punishment, ruling out the

[58]In order to minimize repetitions, when talking about decentralized punishment I will employ the terms "third-party" and "bystander" interchangeably throughout the paper.

effects produced by self-selecting into the same group (contextual interactions) or by sharing common individual characteristics (correlated effects). Focusing on preference interaction means that I study how the utility of a bystander is affected by information about other third-parties' punishment choices when this information does not modify her choice set and payoff complementarity between bystanders is excluded.

Scholars report field and experimental evidence that social influence is a major determinant of human behavior in a variety of settings characterized by important economic consequences, like teenage pregnancy (Akerlof et al., 1996), obesity (Christakis and Fowler, 2007), judicial voting patterns (Sunstein et al., 2006), investment strategies (Hirshleifer and Hong Teoh, 2003), tax evasion (Fortin et al., 2007; Galbiati and Zanella, 2012) and other criminal activities (Glaeser et al., 1996). However, empirically estimating social influence effects on decentralized third-party punishment presents two major identification problems. On the one hand, it requires to rule out problems of self-selection and counfounding factors like correlated effects or the possibility that third parties' material payoff is modified as a consequence of the exposure to social information. On the other hand, in most societies punishment of rule violators is carried out by a centralized system based on codified legal rules that coexists and sometimes overlaps with a decentralized system based on informal norms (Akerlof, 1989; Cooter, 1998). As a consequence, it is often impossible to isolate the effects of social influence on decentralized punishment behavior analyzing field data. Therefore, I exploit the advantage that the controlled environment of a laboratory experiment offers in order to rule out self-selection problems and disentangle the effects of social influence from those of possible counfounding factors.

Furthermore, Deutsch and Gerard (1955) suggest that social influence affects the behavior of an individual agent through two possible channels. On the one hand, under "informational" social influence an agent derives utility from doing what is the right action. Therefore, the agent's behavior is influenced

by information received about peers' choices because of an update of her own beliefs regarding what is the correct thing to do. On the other hand, under "normative" social influence an agent derives (dis-) utility from being (dis-) liked by her peers. Therefore the agent's behavior is influenced by information regarding peers' choices because of the utility gain derived from being liked or the disutility coming from being negatively judged by them. Disentangling informational and normative social influence is important for policy purposes because, while the former has persistent effects on individual behavior, the effects of the latter is less robust and limited in time (Cason and Mui, 1998).

Therefore, in this chapter I address the following questions: is social influence a major driver of third-party punishment? Does social influence play a role in bystanders' punishment decision through the channel of normative or informational influence?

Results of my experiment show that social influence is an important determinant of third-party punishment. Moreover, I find that bystanders engaging in a high level of punishment are affected by social influence the most. I also find that information about peers' behavior influences individual choices the most when the difference between a bystander's beliefs regarding peers' punishment and actual peers' punishment is large. Finally, I find that some subjects respond to normative social influence but not to informational social influence.

I proceed in this way. In the next section I provide a literature review. In section 4.3 the experimental design is presented. Section 4.4 specifies the theoretical framework and the hypotheses I test. Section 4.5 presents the experiment results and section 4.6 discusses my findings, suggests possible directions for future research and states the conclusions.

4.2. Literature Review

Recognizing the importance of punishment in the societal framework scholars have devoted great attention to the topic. Contributions in the early liter-

ature were mostly concerned with second-party punishment (SPP) or forms of centralized third-party punishment (TPP). The milestone of the literature in law and economics could be considered the work of Becker (1968a). The author analyzes the decision of a potential criminal to violate the law in the framework of individual utility maximization, arguing that the criminal act would be carried out only if its expected benefits exceed the expected costs. Therefore, according to Becker's argument, an increase in punishment level and probability of being punished associated to criminal activities would result in the reduction of crime rate. Subsequent contributions extend Becker's original idea to the frameworks of regulation (Bose, 1995) and tax evasion (Slemrod and Yitzhaki, 2002, for a survey).

In the field of experimental economics, the seminal work by Güth et al. (1982) introduces the concept of "irrational punishment" in the context of the so called Ultimatum Game. In this game a receiver has to either accept or reject the share of an amount of money offered by a Proposer. If the offer is rejected, the receiver earns nothing but nullifies at the same time also the earnings of the Proposer. Contrary to game-theoretical prediction, the evidence shows that agents often prefer, at the price of leaving the offer of the Proposer on the table, punishing by rejecting positive offers which she regarded as unfair.

Since Guth's contribution an extensive investigation of the determinants and characteristics of SPP has been conducted. Among others, I mention the contributions of Ostrom et al. (1992), Gächter and Fehr (2000) and Fehr and Gächter (2002) that analyze costly punishment in commons and public goods setting. These studies find that the presence of a punishment mechanism substantially improves cooperation levels.

Subsequent articles further investigate the characteristic of SPP, suggesting that it follows the law of demand (Carpenter, 2007) and it is driven more by an emotional satisfaction than by a rational need for justice (Casari and Luini, 2009), eventually leading to degeneration in riots and resources wasting

(Nikiforakis, 2008).

Despite some pioneering contributions, (Axelrod, 1986; Bendor and Mookherjee, 1990; Ellickson, 1999), it is instead only starting from the last decade that scholars begin to investigate decentralized TPP. The groundbreaking articles are due to Fehr and Gächter (2002); Fehr et al. (2003); Fehr and Fischbacher (2004b). The authors show in laboratory experiments that third parties voluntary incur costs in order to punish norm violations and that the amount of punishment increases with perceived unfairness. Subsequent works by Shinada et al. (2004) and Bernhard et al. (2006) report evidence that humans in their punishment decisions are subject to "parochial altruism", tending to punish the rule-breakers more when he belongs to the same reference group than when he is an outsider. Okimoto and Wenzel (2011) confirm these findings showing that intra-group status influences TPP even if only symbolic and limited to the context of an anonymous laboratory experiment. Moreover Lieberman and Linke (2007) show that social categories have a significant effect on the level of TPP provided and Hoff et al. (2011) find that the legacy of cast culture in India influences norm enforcement, determining less TPP within the casts considered at the bottom of the society. More recently, Coffman (2011) shows that intermediation processes reduce both TPP and rewards and Lewisch et al. (2011) suggest that TPP suffers the free-riding problem when more than a single potential bystander is present.

In a cross-cultural study among 15 small scale societies, Henrich et al. (2006) find a significant variability on the level of TPP provided. Attempting to account for these differences, Marlowe et al. (2008) show that societies characterized by complex organizations and subject to frequent market interactions engage in higher level of TPP compared to less articulated ones. As a consequence, the authors argue that institutions and social structures play a fundamental role in shaping our preferences for punishment. However, possibly in contrast with Marlowe's findings, Mathew and Boyd (2011) in a

following work show that third-party punishment could sustain large scale cooperation among African nomadic tribes during warfare period, suggesting that further studies are necessary in this area of research.

Contributions in applied psychology investigating the determinants of decentralized TPP have also flourished in the last decades. Kurzban et al. (2007) in a laboratory experiment find that subjects increase punishment when observers are present, arguing that TPP is influenced by the so called "audience effect". Subsequent works confirm that anonymity has a causal effect on TPP (Piazza and Bering, 2008), suggesting that the third party decision to sanction wrongdoers is influenced by a cost-dependent reputation effect (Nelissen, 2008) and by emotions (Nelissen and Zeelenberg, 2009). Moreover Lotz et al. (2011) suggest that differences in the level of third-parties punishment provided within a group of agents could be explained by heterogeneity in bystanders' "justice sensitivity". Finally, a promising branch of research aims at explaining TPP through the investigation of the biological mechanisms governing the human brain (Seymour et al., 2007; Buckholtz et al., 2008) or the behavior of other animal species (Raihani et al., 2010).

While to the best of our knowledge there are no contributions linking social influence effects and decentralized third-party punishment, nevertheless the possibility that agents' behavior is influenced by peers has since a long time been the object of interest for social scientists. Depending on the field of study and the context of the research, this behavior is called "social influence", "neighborhood effect", "taste for conformism", "imitation" or "herd behavior". Starting from the pioneering work of Asch (1951, 1956), contributions in experimental psychology show how individuals tend to modify and distort self-judgments under the influence of group pressure, culture influence and taste for conformism (for a survey see Bond and Smith, 1996).

Economists have been mostly interested in the implications of social influence effects for the functioning mechanisms of financial markets. Indeed, most of the contributions focus on the process of information acquisition in

investment strategies (Cooper and Rege, 2008; Devenow and Welch, 1996; Scharfstein and Stein, 1990; for a survey see Hirshleifer and Hong Teoh, 2003). Also, economic scholars investigated the effects of social influence on the labor market. Studies report that peer pressure influences labor productivity (Falk and Ichino, 2006; Mas and Moretti, 2009) and that social networks characterized by an elevated percentage of unemployed individuals could generate social norms perpetuating unemployment (Akerlof, 1980; Topa, 2001). Moreover, reporting results of laboratory experiments, Falk and Fischbacher (2002) argue that social influence is a major driver of criminal behavior and Falk et al. (2010) and Krupka and Weber (2009) find that social influence plays a role in determining pro-social behavior.

Finally, I signal a series of recent policy interventions that exploits social influence effects in order to achieve welfare-improving results (for a discussion of further similar policies see also Thaler and Sunstein, 2008). The first framework of intervention is related to tax compliance. I already mentioned in the previous chapter that for an individual the likelihood to engage in tax evasion is affected by her peers' rate of tax evasion . However, social influence properly combined with framing effect[59] might also help in increasing tax compliance. Indeed,Coleman (1996) reports the result of a field experiment conducted in Minnesota with the objective to find a costless strategy to increase tax compliance. In the experiment, a letter is sent to taxpayers by the tax authority a few days before the annual tax file deadline. The letter could contain different information according to different treatments: a reminder to the civic obligation to pay taxes, information regarding the procedure to follow, a simple reminder of the deadline or information regarding the number of taxpayers that have already complied with tax payment (a percentage above 90%, that a survey analysis reveals people typically underestimated). This last treatment turns out to be the only treatment

[59]See the Introduction chapter for a discussion of the framing effect.

registering a statistically significant increase in the number of compliant tax payers.

Another example of behavioral public policy based on social influence is an intervention that encourages socializing nondrinking. Perkins et al. (2010) report results from a field experiment run in Montana, where surveys revealed that college students systematically overestimate the fraction of peers consuming alchoolic beverages. In the experiment, a random sample of college students were exposed to the (true) information that the overwhelming majority of the people in the state and of the students on campus consumes moderate quantities of alchol. Results show that those exposed to this information decrease the consumption of alchool compared to peers in the control group.

The last example I report is described in Cialdini (1993) and concerns a field experiment run in the Petrified National Forest, Arizona. Apparently, visitors tend to take home petrified fossils as a souvenir, a behavior that over the years created serious concerns about the preservation of the park. Signs along the park trails ask people not to take samples away. However, Cialdini found that when the request on the sign is framed as an injunctive norm ("Please do not remove fossils from the park in order to preserve the Petrified Forest") people are significatively more compliant than when the request conveys information about other visitors' unlawful behavior ("Many past visitors have removed the petrified wood from the park, changing the state of the Petrified Forest"). This finding confirm that social influence, either for good or bad, represents a major driver of human behavior.

4.3. Experimental Design

The Game. I conduct a variant of Fehr and Fischbacher (2004) dictator game with TPP. Following Cox et al. (2007) and Swope et al. (2008), in the game a dictator has the possibility to take from a passive receiver some or all of the experimental monetary units (tokens) of the initial endowment provided

by the experimenter. The game has 3 possible roles: receiver (Participant A), dictator (Participant B) and Third-party (Participant C).
The game has two periods. Each period of the game is divided in two stages. At the beginning of each period, each participant is endowed with 30 tokens by the experimenter. In the first stage of each period, Participant B has the possibility to take from 0 up to 30 tokens (in multiples of 5) from A. Participants A cannot undertake any action during the game.
In the second stage of each period, Participant C has the opportunity to impose a costly punishment upon B. Specifically, C could use up to 20 units of her initial endowment to reduce B's payoff. For each token used by C, the payoff of Participant B is reduced by 4 tokens. Participants C specify how many tokens they use in order to reduce B's payoff for each possible action choosen by B (strategy method). The tokens C uses for punishment in one period and the consequent reduction of B's payoff have no effect on the payoff of player A. Agents have full information regarding the rules of the game.
Before the game starts, participants' beliefs about the average punishment choices of the peers are elicited. To do so, I use an incentivized coordination game similar to Krupka and Weber (2013). I refer to this part of the experiment as the "Beliefs elicitation game". I present to participants a hypothetical situation identical to the game described above. I ask each participant to indicate, for each of the seven possible actions of B, the number of tokens [0; 20] that in their opinion C would use to punish B. I explain that, once each participant present in the laboratory has provided her answers, the computer selects one of the seven possible actions of B. For the selected action, a participant earns 40 tokens if the number she indicated is equal, bigger or smaller by one unit to the average number indicated by all the participants in the experimental session. Therefore, in this part of the experiment participants have incentives to reveal their true beliefs regarding

peers' choices of punishment[60].

Treatments. I propose two effect treatments (INFORMATIONAL and NORMATIVE) and a control treatment (CONTROL). The elicitation of beliefs and the first period of the game are identical in all treatments. Specifically, the amount of tokens B decides to take from A in the first stage is not immediately observed by C. Instead in stage 2 of the first period C's decisions of punishment are elicited employing the "strategy method": for each possible action of B, C states his decision of punishment. Participants are informed that only the punishment decision corresponding to the actual choice made by B determines payoffs. The punishment tokens used by C in correspondence to the other possible choices of B do not have payoff consequences. First period earnings and choices of peers are not revealed to participants at the end of the first period.

At the beginning of the second period, participants' endowments are restored to the initial level. Earnings of the first period are independent from those of the second. The first stage of the second period is identical to the first stage of the previous one: B is endowed with the same amount of tokens and may take part or all of A's endowment.

Also in the second stage of the second period, C has to indicate the level of

[60]One may argue that eliciting subjects' beliefs regarding average peers' punishment in the first part of the experiment might influence subjects' choice of punishment in later parts. Indeed, it is possible that individuals anchors their punishment choice to the expected average punishment. I considered this point carefully in designing the experiment. However, on the one hand there is an unavoidable trade-off between eliciting subjects' beliefs and facing the risk of an anchoring effect. Given the importance that subjects' beliefs have in my model, I could not avoid this stage. On the other hand, this concern would be jusified for experiments that does not involve an incentive mechanism, while is definitely less worrisome for my experiment that is based on the standards of experimental economics. In fact, in my experiment subjects are paid according to their choices. Therefore, for a subject interested in maximizing monetary earnings, the incentive schemes proposed guarantees that in the second part of the experiment any anchoring effect would be eliminated or reduced.

punishment inflicted for each of the 7 possible actions of B (take 0 from A; take 5 from A;...take 20 from A). The difference between treatments consists in the kind and amount of information disclosed to participants C before the punishment choices. In the INFORMATIONAL treatment, each participant C receives information about the average number of tokens used to punish B in the first period by the participants C taking part to the experimental session.

In the NORMATIVE treatment, each participant C receives the same information of INFORMATIONAL. However, she is additionally informed that her punishment decisions of the second period will be revealed to 5 peers randomly selected among the experiment participants. After observing these choices, the 5 peers vote for sending an emoticon that will appear to the screen of the participant C. The 5 peers could vote for a smiling emoticon or a sad emoticon. If the majority vote for a smiling emoticon, on player C's monitor will appear a smiling emoticon. A sad emoticon will appear on the screen otherwise. Participants are informed that the emoticon received has no effect on earnings and that it disappears after one minute.

In the CONTROL treatment, no relevant information about participants punishment choices is disclosed in the second period. However, I have to rule out the possibility that a change in punishment behavior between periods is driven by factors other than the exposure to social influence. One possible confounding factor is subjects' experience that increases between periods. Another possible confounding factor is that processing new information imposes a cognitive effort to subjects. Hence these factors, not the social content of the information received, could be responsible for a modification of the punishment choice. In fact, in the INFORMATIONAL and NORMATIVE treatments subjects have to process some sort of information, and there is evidence that individuals exposed to a cognitive load tend to modify their behavior (for discussion on this point see Cason and Mui, 1998).

In order to rule out these confounding factors and isolate social influence effects, I then expose the CONTROL group to some social irrelevant information. Specifically, I ask at the beginning of the session to each participant her day of birth $\in$ [1; 31] and I take the average. Since I do not ask nor I report them neither the year nor the month of birth, reporting this measure does not convey any relevant social information. However, in this way participants in CONTROL are affected by the same cognitive burden of participants in TREATED and the only difference lies in the exposure to relevant social information .

Endowments, Choice Sets and Payoffs.
Participants's initial endowment in each period of the Dictator game is 30 tokens. The initial endowment is restored at the beginning of each period. Earnings of the first period are independent from those of the second one. In each period, first B decides how many tokens to take from A. B could take from 0 up to 30 tokens in multiples of 5 from A. Then, C decides how many tokens to use for punishing B. C could use from 0 up to 20 tokens of his initial endowment. For participants C, the cost of reducing the payoff of B of 4 tokens is 1 token. Only punishment of integer tokens is allowed. Participants are informed that eventual negative earnings would be deducted from the participation fee. C has to take 7 punishment decisions. In fact, C does not observe the actual choice of player B. Instead, C reports his punishment choice for each of the 7 possible B's actions.
In all treatments, the per period payoffs are calculated as:

- Π_A = 30 - t
- Π_B = 30 + t - 4*p
- Π_C = 30 - p

where:

- t = tokens taken by B from A
- p = punishment tokens used by player C

Results and earnings of the beliefs elicitation game and of the first period of the dicator game are not revealed to subjects until the end of the experiment. In order to calculate individual earnings, participants are randomly divided in groups of 3. Each group is composed of one participant A, one B and one participant C. The final payment for each group follows this procedure:

- In the dictator game, after the second period is concluded, one of the two periods is randomly selected. This period will be called the "payment period".
- For each participant, earnings relative to the payment period are added to earning collected in the beliefs elicitation game. Earnings from the non-selected period are not paid out.
- A 5 euro participation fee is added to total payments.

Given the experimental design, I am able to isolate the effect of *endogenous interactions* in the form of *preferences interactions* on TPP. This is possible because I randomly select and assign participants to roles, I exclude payoff complementarity among third-parties and I explicitly present them a choice set that remains unchanged throughout the experiment. I also feel confident that my design rules out the possibility that individuals are influenced by "epistemic norms" (Hetcher, 2004). Epistemic norms emerge when agents experience scarcity of information and so just "follow the crowd" in taking a decision. To be precise, epistemic norms are not even norms but conventions motivated by simple self-interest (Kahan, 1997). Independently from how these norms are named, I believe they are not playing a role in the experiment, since every player has perfect information about possible actions in his choice set, and payoff complementarity and strategic actions are ruled out

by design.

The Procedure. The experiment was programmed using the software z-Tree (Fischbacher, 2007). Every session was conducted at the Bologna Laboratory for Experimental Social Sciences at the University of Bologna, Italy, between November and December 2013. Participants were for the vast majority graduate and undergraduate students of the University of Bologna, plus some private citizens, and were recruited through the online system ORSEE (Greiner, 2004). In each session participants were split into 5 groups of 3 subjects[61]. Overall, 9 sessions were run, 3 for each treatment, that results in a total of 142 participants (56% female).

In each session, before each of the three parts of the experiment (elicitation of beliefs, first period punishment and second period punishment), a printed copy of the instructions was distributed and read aloud by the experimenter[62]. Participants had additional images and tables summarizing the instructions on their computer screen. Information regarding payoff functions and rules of the game was common knowledge. Participants had the possibility to ask questions before the experiment started.

At the end of each session participants completed a brief socio-demographic questionnaire[63]. Each participant took part in one session only. Peers' identities were maintained unknown even after the end of the experiment. In order to guarantee anonymity, participants were individually and privately paid after the experiment finished. No communication among participants was allowed.

[61] In one session of the INFORMATIONAL treatment there were only 4 groups, for a total of 12 participants.

[62] Original instructions are in Italian and are available upon request. A copy of the instructions for the NORMATIVE treatment translated in English is included in Appendix B.

[63] In one session of the INFORMATIONAL treatment subjects' socio-demographic characteristic were not recorded due to a technical problem.

The part of the session concerning beliefs elicitation and treatments lasted around 20 minutes. However, due to the impossibility of learning throughout periods and the limited number of decisions each participant had to take, I was concerned about the possibility that instructions were not fully understood. In order to minimize this possibility, I adopt special care in writing detailed instructions and providing multiple examples, and I also asked subjects to correctly answer control questions before proceeding with each part of the experiment. As a result, each experimental session lasted in total about 45 minutes. Tokens where converted into euros at a rate of 5 tokens for 1 euro. Subjects earned on average 11 euros for the experimental session.

4.4. Hypotheses

Following the customary assumptions, the predictions of the game outcomes are straightforward. Agents' utility is an increasing function of individual wealth and agents are individual payoff maximizer. Hence, in any treatment, no punishment should be observed, since the payoff-maximizing strategy for third-parties is to punish nothing and keep the initial endowment. Anticipating the absence of punishment, dictators should take all the tokens from receivers.

However past dictator game experiments have shown two behavioral regularities. On the one hand, even in games where the dictator faces no threat of punishment, positive amounts of tokens are transferred (in our setting: are left) to the receiver. On the other hand, third-parties engage in costly punishment for dictator's levels of transfer (in our setting: for dictator's levels of taking) perceived as unfair. In this study I am interested in verifying how, given the action of a dictator, the punishment choices of other third-parties affects the utility that a bystander derives from punishing the dictator.

Consider the choice of a third-party i to use p tokens of her initial endowment in order to punish a dictator that takes z tokens from a passive receiver. Third-parties' individual utility is an increasing function of the final monetary earnings x. Moreover, given a dictator's action, third-parties have some

inherent preferences p_z^k for the amount of tokens she wants to use for punishment. $p_{i,z}^k$ could be interpreted as reflecting the individual sense of justice of the third-party i. If a third-party chooses to punish the dictator a quantity different from her inherent preference, she has to bear a cost s that increases when the absolute difference between p^k and the p increases.
Furthermore, third-parties have some beliefs $E(\bar{p})$ regarding the average amount of tokens that the other bystanders will use for punishing dictators. A third-party incurs a cost c for punishing a quantity of tokens different from $E(\bar{p})$, and this cost becomes larger when the absolute difference between individual punishment and the average punishment of the peers increases. c incorporates both the costs imposed by the other bystanders observing the third-party that deviates from the average punishment and the disutility the third-party experiences in not conforming with the peers' behavior independently from the fact that her action is observed.
Therefore, in her punishment decisions a third-party maximizes individual utility taking into account the cost of using tokens for punishing a dictator and so reducing her monetary payoff, the cost for deviating from her inherent preference for punishment and the cost of not conforming to the peers' average punishment:

$$\begin{aligned} \max_{p_{i,z,t}} \quad & U_{i,z,t} = x_{i,z,t} - (s(E_i(\bar{p}_{z,t}) - p_{i,z,t})^2 + c(p_{i,z}^k - p_{i,z,t})^2) \\ \text{s. t.} \quad & y_i = x_{i,z,t} + p_{i,z,t} \end{aligned} \tag{24}$$

Where y is third-party's initial endowment. Assuming an interior solution exists, equation (1) generates the following first order condition:

$$p_{i,z,t}^* = \frac{sE_i(\bar{p}_{z,t}) + cp_{i,z}^k - 1}{s + c} \tag{25}$$

Therefore, according to our model of social influence the optimal punishment choice of a third-party is an increasing function of the expected punishment chosen by her peers. Furthermore, the higher the cost s of not conforming

to other bystanders' average punishment relative to the cost c of deviating from inherent preferences, the higher will be for a third-party the tendency to conform to the peers' average punishment. Allowing for concavity of the agents' utility function, the intutition of the previous results will still work.
In order to test my predictions, as a first step I verify if there is a positive association between a third-party's beliefs regarding peers' average punishment and her first period punishment. As a second step, I then investigate how participants modify their punishment choices between the first and the second period. Assume that third-parties in TREATED revise beliefs about average peers' punishment substituting their initial priors with the actual punishment level revealed to them after the first period, hence $E_i(\bar{p}_{z,t}) = \bar{p}_{z,t-1}$. The punishment variation across periods is given by:

$$(p^*_{i,z,2} - p^*_{i,z,1}) = \frac{s(\bar{p}_{z,1} - E_i(\bar{p}_{z,1}))}{s + c} \tag{26}$$

For the moment, focus only on the distinction between participants in CONTROL and the other participants grouped together, that I call TREATED. Third-parties in CONTROL are not exposed to socially relevant information between period 1 and 2. Instead, bystanders in TREATED are exposed to information that may induce them to update their initial beliefs regarding peers' average punishment and so influence their second period punishment decision. Therefore, if social influence has an effect on third-party punishment decision, I expect it to be more likely that participants in TREATED modify their punishment decisions in the passage between the first and second period punishment as compared to participants in CONTROL.
Thus, according to our model revealing to a bystander her peers' average punishment may trigger a change in her second punishment decision as a consequence of a beliefs updating process. Specifically, for a bystander the likelihood to change punishment decision in the second period increases when the absolute difference between her beliefs regarding peers' average punishment and the actual average punishment of the first period is large. There-

fore, I test the following hypothesis:

1. Zero Social Influence hypothesis: *In the first period, punishment decisions of bystanders are not influenced by their beliefs regarding peers' average punishment. Moreover, bystanders in TREATED are as likely as bystanders in CONTROL to modify their initial punishment decisions.*

Second, I want to identify who are the bystanders more responsive to social influence. Third-parties deciding to use tokens for punishing a dictator are reducing their final monetary payments. Hence, every time I observe a bystander punishing a positive amount, according to our model I infer that $sE(\bar{p}) + cp^k - 1$ is positive. This could mean that the bystander has inherent preferences for punishing a positive amount (p^k>0) and at the same time she attaches a positive weight to this component of the utility function (c>0). However, it is also possible that the bystander attaches a positive weight to the social component of the utility function (s>0) and she expects peers to punish on average a positive amount of tokens ($E(\bar{p}_{z,t})$>0)[64]. If this last possibility is true, the higher a bystander's punishment in the first period the more she attaches weight to the social component of the utility function and so the more likely she is to modify the second period punishment decision.
Now consider the difference between first and second period punishment of a bystander. Inherent preferences for punishment are stable, so they do not play a role in the decision to eventually modify punishment choice. Instead, according to the prediction of my model, the larger is s for a bystander, the more she responds to the social information regarding peers' punishment. Therefore, holding constant $E(\bar{p}_{z,t})$ - $\bar{p}_{z,t-1}$, I expect that the more a bystander punished in the first period, the more she is likely to revise her

[64]Of course, it is possible that what it is observed is a combination of this two possibilities.

punishment decisions in the second period.
Moreover, I also consider the difference between a bystander's beliefs regarding peers' average punishment and her first period punishment. In the first period, a bystander could punish an amount different from her beliefs regarding peers' average punishment because she only cares about her monetary payoff or because her inherent preference for punishment differs from the expected average punishment and the cost s of non conforming to peers' average punishment is small compared to the cost c of non following inherent preferences. In both cases, the choice of the bystander reveals that in her punishment decisions she is little influenced by peers' behavior. As a consequence, I expect the more a bystander punishes in the first period a quantity different from her beliefs about peers' average punishment, the less she will be responsive to social influence.

2. Differential Social Influence hypothesis: *Third-parties that engage in high punishment in the first period are the most responsive to social influence. Conversely, the higher the absolute difference between a bystanders' beliefs regarding average peers' punishment and first period individual punishment, the lower the bystander likelihood to modify punishment choices in the second period.*

Finally, I investigate the psychological mechanisms triggering social influence. In our experiment, I give bystanders in NORMATIVE and INFORMATIONAL the same information about peers' punishment. However, in the INFORMATIONAL treatment, the second period choices of the third-party are not observable ex post by other participants. As a consequence, in the INFORMATIONAL treatment a third-party has no incentives to conform to peers' punishment choices if her only goal is being liked by them. Hence, a bystander would eventually modify his punishment strategy only if informational social influence is at work.

On the other hand, in the NORMATIVE treatment a bystander is aware that her punishment decisions of the second period will be observed by peers and that they will express a judgement regarding those choices. As a consequence, for bystanders in NORMATIVE the cost s of not conforming to the average peers' punishment has been modified in the passage between first and second treatment since I added a normative social influence component. Hence, if some bystanders are responsive to normative but not to informational influence, the NORMATIVE treatment will show social influence effects different from those resulting from the INFORMATIONAL treatment. The difference in the way bystanders modify their punishment decisions between NORMATIVE and INFORMATIONAL treatments isolates the effect of normative social influence on third-party punishment.

3. Equivalence of Normative and Informational Influence hypothesis: *Social influence effects on third-party punishment are the same for subjects exposed to informational and normative influence.*

4.5. Results

Table 5 reports summary statistics relative to my data[65]. Dictators leave approximatively 36% of receivers' endowment. This finding is consistent with results from other comparable experiments where dictators have to take tokens from the endowment of a passive receiver (List, 2007; Krupka and Weber, 2013)[66].

[65] Additional summary statistics where I consider separately the seven possible punishment choices in each period, are reported in Tables B.12 and B.13 in Appendix B.

[66] In the classical dictator game without punishment a dictator has the possibility to give part of his endowment to a passive receiver. In a meta-study, Engel (2011) found that on average dictators give roughly 25% of their endowment to the receiver. However, in my design dictators has to *take* money from receivers' endowment instead of *giving* them. This difference and the possibility of being punished that characterizes my design

Table 5: Summary Statistics

Treatment	male	age	dictatorTake	Beliefs	PunishPer1	PunishPer2
Control						
(Mean)	.33	24.68	18.38	5.08	3.33	3.54
(Median)	0	25	20	4.29	2.71	2.71
(SD)	.48	2.76	11.19	3.71	3.42	3.82
Normative						
	.58	26.18	18.28	4.80	3.30	3.03
	1	25	20	4.57	3	2.29
	.50	5.38	11.40	3.40	3.34	3.37
Informational						
	.37	25.15	17.32	5.41	4.15	3.81
	0	25	16.25	4.86	4.29	3.86
	.49	3.72	10.85	3.67	3.71	3.90
Total						
	.44	25.37	18.01	5.09	3.58	3.45
	0	25	20	4.71	3.43	3.07
	.50	4.17	11.08	3.57	3.48	3.69

On average bystanders punish approximately 3.5 tokens, decreasing punishment amount in the second period. When the dictator takes all the money from the receiver, third parties spend approximately 6 tokens in punishment. Average punishment then progressively declines, reaching virtually 0, when levels of dictators' taking decrease. Also this result is consistent with previous findings on third-party punishment in dictator games (Fehr and Fischbacher, 2004b). However, if I consider the 3 treatments separately,

are likely to explain the slightly more fair allocation I registered compared to the standard dictator game (on this point, see also Krupka and Weber, 2013).

I see that in CONTROL punishment slightly increases in the second period, while in both NORMATIVE and INFORMATIONAL it decreases. Considering third-parties' beliefs regarding peers' punishment behavior, beliefs are on average are higher than actual punishment.

I proceed considering, for each bystander in a single punishment period, the average of her seven punishment choices corresponding to different levels of dictator taking. I compare the cumulative distribution of this measure in TREATED and CONTROL. Figures 4 and 5 report the cumulative distribution functions of punishment respectively in period 1 and 2. In the first period, the cumulative punishment choice distribution in CONTROL exceeds the distribution in TREATED for any possible punishment level. However, a Kolmorgorov-Smirnov test cannot reject the null hypothesis that the distributions are equivalent. In the second period instead, the cumulative punishment choice distribution of TREATMENT exceeds the distribution of CONTROL for some punishment levels greater than 5. However, also in the second period a Kolmorgorov-Smirnov test cannot reject the equivalence of the two distributions, nor the samples means are statistically different (t-test two sided p-value 66%). Now I test my hypotheses.

4.5.1. Zero Social Influence Hypothesis

I start by investigating if bystanders punish in the first period according to their beliefs regarding the average punishment they expect peers' will use. As a first step, I test if there is a significant difference between the punishment used by a bystander and her beliefs about peers' average punishment. I conduct a t-test comparing the two averages under the null hypothesis that they are the same. Third-parties punish on average 3.6 tokens in the first period while their beliefs about peers' average punishment is 5.1 tokens.

Results of the t-test reject our hypothesis and indicate that bystanders in the first period punish significantly less than what they think peers on average will do (t-test two-tails, p-value $< 1\%$).

I want to verify if this result is driven by those third-parties that during the

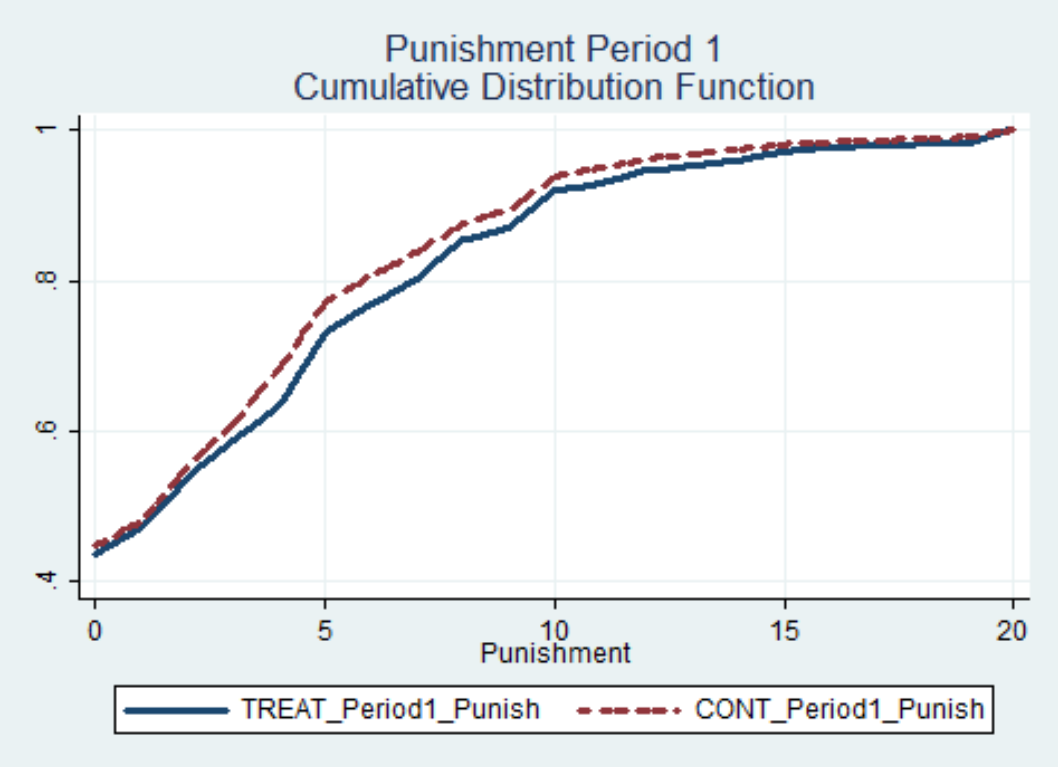


Figure 4: Punishment Period 1 Cumulative Distribution Function

experiment punish always 0 (I name them "selfish"). Excluding selfish punishers from the sample, bystanders' beliefs about peers' punishment remain higher than the punishment they provide (6.0 versus 5.4 tokens), however the difference is not statistically significant (p-value 12%). This results suggest that selfish subjects are responsible for the aforementioned gap.

I continue the analysis regressing the quantity of punishment tokens a bystander uses in the first period with her beliefs regarding peers' average punishment and a set of socio-demographic characteristics. Results are reported in Table 6[67]. The variable $Beliefs\_Punish$ indicates bystanders'

[67]Table C.14 in Appendix C reports a description of each variable employed.

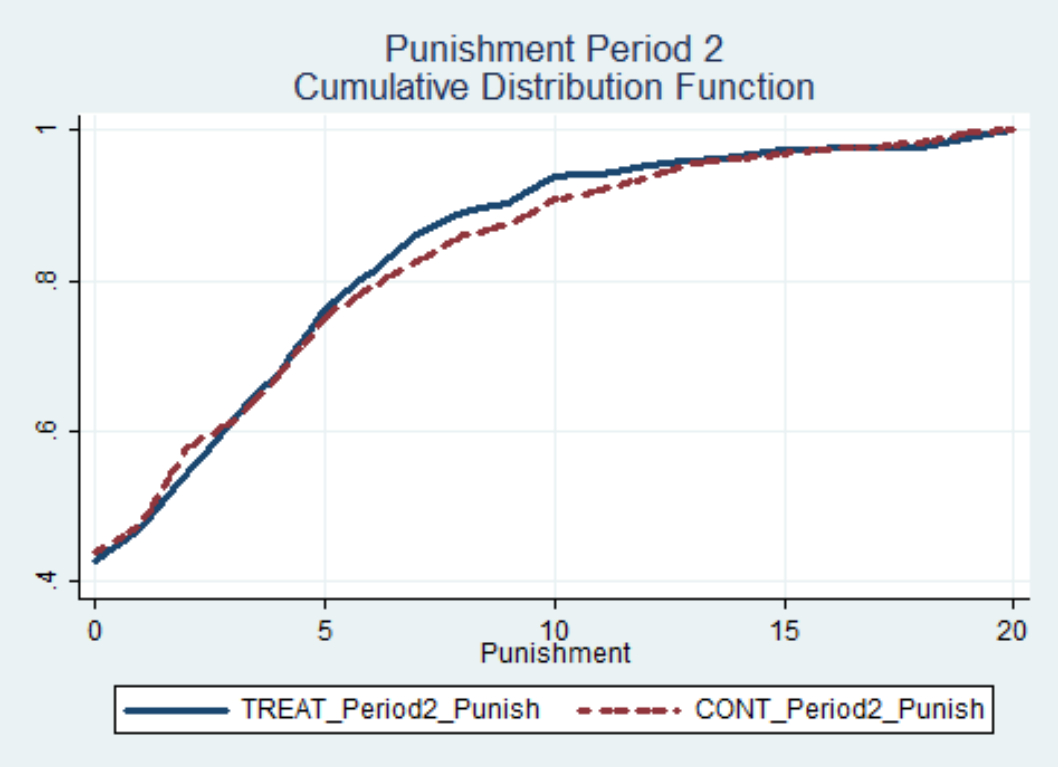


Figure 5: Punishment Period 2 Cumulative Distribution Function

beliefs about peers average punishment. The coefficient is positive and significant at the 1% level in all model specifications. According to the model estimation, bystanders spend an additional 0.4 token for every unit of increase in expected average peers' punishment. Hence, data suggest that third-parties are influenced in their first period punishment decisions by beliefs about peers' punishment. This finding goes against the Zero Social Influence hypothesis.

I proceed in the analysis verifying how third-parties in CONTROL and in TREATED modify punishment choices between the first and the second pe-

Table 6: Determinants First Period Punishment

	(1)	(2)
Beliefs	0.405***	0.432***
	(0.05)	(0.06)
male	-0.515	-0.407
	(0.58)	(0.62)
age	-0.078	-0.047
	(0.06)	(0.07)
degree	-0.842**	-1.435***
	(0.41)	(0.45)
worker	0.759	0.860
	(0.74)	(0.80)
social	0.378	0.399
	(0.73)	(0.85)
arts	0.234	0.021
	(1.06)	(1.28)
field_other	0.409	-0.131
	(0.63)	(0.75)
risk	0.138	0.037
	(0.10)	(0.11)
logic	-0.305	0.057
	(0.40)	(0.47)
impulsivity	-0.636**	-0.687**
	(0.27)	(0.32)
Instruction	0.000	-0.000
	(0.00)	(0.00)
DictatorTake	-0.007	0.026
	(0.01)	(0.02)
_cons	6.246***	7.889***
	(1.90)	(2.04)
N	924	630
R^2	0.281	0.275
BIC	5203.7	3602.6

Notes: OLS regression: dep. var. $Strat\_Punish$, SE clustered by subject. Significance levels: * p<0.10, ** p<0.05, *** p<0.01

riod. As a first step, I sort bystanders into two main categories: those who never change punishment decisions across periods and those who change at least once. In the TREATED group 53 subjects (61%) change at least one punishment decision between periods, while in the CONTROL treatment 24 subjects (53%) do so. This difference is not statistically significant. If I repeat the same test excluding selfish bystanders, it turns out that in TREATED 87% and in CONTROL 80% of third-parties change punishment decisions at least once. However, also in this case the difference is not statistically significant.

I also verify how many times on average each punisher changes decision across periods. In TREATED bystanders change decision 2.5 times, while in CONTROL they change 2.1 times. This difference is not statistically significant and it remains roughly unchangend even if I exclude selfish bystanders. Therefore, these results do not provide evidence against the Zero Social Influence hypothesis. The result seems to be driven by the high percentage of participants (53%) in the CONTROL group that modifies punishment choices at least once, even if they did not receive any relevant social information.

As a second step, I test if there is a difference in the likelihood that participants in CONTROL and TREATMENT change punishment decisions. I create the dummy variable $DummyP1p0$ that takes the value 1 when punishment in the second period differs from punish in the first one and 0 otherwise. I implement a logistic regression to estimate the likelihood of changing punishment choice across periods. Results of the model are presented in Table 7. The dummy $TREATED$ equals 1 for participants in NORMATIVE and INFORMATIONAL. The coefficient of the dummy is positive and statistically significant in any of the model specifications[68]. Therefore, I conclude that

[68]Model 2 differs from Model 1 because it excludes selfish participants. Model 3 adds the control variables $Strat\_Punish$, indicating punishment exerted in the first period, and $Abs0Belifs$, reporting the absolute difference between a bystander's punishment in the first period and her beliefs regarding peers' average punishment. Model 4 excludes selfish participants from the sample.

the results of the logistic regression do not support the Zero Social Influence hypothesis and indicate that participants in TREATED modify punishment decisions across periods more often than those in CONTROL.

As a third step, I investigated how third-parties modify their punishment choices. In CONTROL bystanders reduce punishment in the second period 48 times (15%), increase 49 times (16%) and do not change 218 times (69%). In TREATED, bystanders reduce punishment 140 times (23%), increase 93 (15%) and do not change 376 times (62%). These choices result for CONTROL in an average increase in punishment from period 1 to period 2 of 0.21 tokens (from 3.3 to 3.5) and in an average decrease in TREATED of 0.30 tokens (from 3.7 to 3.4). The mean punishment difference across periods is not statistically different in CONTROL and TREATED (p-value 0.16, t-test two-sided).

However, we could expect that bystanders have no reason to punish a dictator when she does not take any amount of money from the receiver. Hence, when the dictator chooses to take 0 from the receiver, I expect little or no punishment both in CONTROL and TREATED. In fact, if we exclude the situations in which the dictator takes 0 from the receiver, the average difference between bystanders' punishment in the two periods is weakly statistically significantly higher in CONTROL versus TREATED (p-value 0.09). Furthermore, if we consider only situations in which the dicator takes half or more of receiver's initial endowment, this difference between CONTROL and TREATMENT becomes significant at the 5% level.

Hence, results of this third set of tests suggest that, at least for situations where dictators subtract positive amounts of tokens from receivers, bystanders exposed to social influence significantly reduce the amount of punishment provided compared to bystanders in CONTROL. These results provide evidence against the Zero Social Influence hypothesis.

Fourth, I test the hypothesis that a large absolute difference between a bystander's beliefs regarding peers' average punishment and actual peers' av-

erage punishment increases the likelihood to modify the initial bystander punishment choice. Third-parties receive information regarding actual peers' punishment in the NORMATIVE and INFORMATIONAL treatments only, so I restrict the analysis to these treatments. I test this hypothesis using a logistic model. Results are reported in Table 7. From models 7 and 8, I can see that the coefficient of the variable $Abs\_BeliefAvgPunish$ is positive as expected, however only weakly significant. The estimations suggest that an increase of one unit in the difference $Abs\_BeliefAvgPunish$ increases on average the probability of modifying second period punishment by 3.5%[69]. This result provides evidence against the Zero Social Influence hypothesis.
Finally, I report some descriptive statistics that account for the direction of punishment deviation between periods[70]. The variable $P1p0$ reports the difference between punishment in period 2 and period 1. Table 8 provides summary statistics of this variable, grouping subjects according to the difference between individual beliefs about average punishment and actual average punishment observed. When beliefs match exactly the average punishment observed ($P1p0\_BAP = 0$), subjects confirm punishment choices in the second period 76% ot times. Instead, when beliefs are larger or smaller than the actual average punishment observed ($P1p0\_BAP >$ and < 0 respectively), subjects confirm first period choice respectively 51% and 73% of times.
Considering how subjects modify their decisions, I see that those observing

[69]I also consider the possibility that a large absolute difference between a bystander punishment in the first period and the average peers' punishment increases the likelihood to change the punishment decision in the second period. I create the variable $Abs\_Signalp0$, reporting the absolute difference between individual punishment in the first period and average peers' punishment. Results of the logistic estimations are reported in model 7 and 8 of Table 7. The coefficient of $Abs\_Signalp0$ is not statisticaly different from 0 in any model specification. As a consequence, I conclude that $Abs\_Signalp0$ has no impact on subjects' likelihood to modify punishment decision.

[70]It would be interesting to test if the difference between agents' punishment choices across periods has the same sign of the difference between beliefs and actual average punishment of the peers. However our data do not allow to distinguish between this hypothesis and a simple regression toward the mean.

Table 7: Probability modify punishment across periods

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
male	-0.219***	-0.258***	-0.168**	-0.215**	-0.098	-0.019	-0.089	-0.017
	(0.08)	(0.09)	(0.07)	(0.09)	(0.12)	(0.14)	(0.12)	(0.14)
risk	0.030*	0.014	0.019	0.008	-0.015	-0.040	-0.020	-0.043*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)
logic	-0.077	-0.033	-0.061	-0.031	0.135	0.242**	0.149	0.246**
	(0.06)	(0.07)	(0.06)	(0.07)	(0.09)	(0.10)	(0.09)	(0.10)
TREATED	0.155**	0.131*	0.124**	0.122*				
	(0.06)	(0.07)	(0.06)	(0.07)				
Strat_Punish			0.040***	0.021**	0.044***	0.025**	0.047***	0.029**
			(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Abs_p0Belifs			-0.010	-0.003	-0.026**	-0.019*	-0.032**	-0.026
			(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Abs_Signalp0							0.002	0.006
							(0.01)	(0.01)
Abs_BelAvgPun							0.024	0.023
							(0.02)	(0.03)
Beliefs			0.015	0.009	0.025**	0.021*	0.015	0.009
			(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Other contr	Y	Y	Y	Y	Y	Y	Y	Y
N	924	630	924	630	609	420	609	420
pseudo R^2	0.092	0.059	0.207	0.095	0.206	0.101	0.226	0.123

Notes: Logistic regression: dep. var. $DummyP1p0$, marginal effect at means, SE clustered by subject. Significance levels: * p<0.10, ** p<0.05, *** p<0.01. Other Controls include: degree worker social arts field_other DictatorTake impulsivity age.

actual average punishment smaller than beliefs reduce on average punishment in the second period of 0.30 tokens. Instead, subjects observing actual average punishment equal to beliefs reduce punishment of 0.14 tokens between the two periods, and those observing average punishment larger than beliefs increase punishment in the second period of 0.27 tokens.

- Conclusion relative to the Zero Social Influence hypothesis: *Bystanders' punishment choices in the first period are positively associated with their*

Table 8: Punishment difference across periods

Treatment	**P1p0**	**P1p0_BAP>0**	**P1p0_BAP=0**	**P1p0_BAP<0**
Informational	294	142	27	125
(mean)	-.33	-1.04	-.11	.42
(sd)	3.28	3.96	.42	2.52
Normative	315	170	60	85
	-.28	-.49	-.15	.06
	2.58	3.09	1.63	1.90
Total	609	312	87	210
	-.30	-.74	-.14	.27
	2.94	3.52	1.37	2.29

Notes: Variable $P1p0$ indicate the difference between punishment in first and second period. P1p0_BAP $>$, $<$, $= 0$ indicate respectively P1p0 when individual beliefs regarding average punishment are $>$, $<$, $=$ actual average punishment.

own beliefs about average peers' punishment. However, I find that on average bystanders punish less than the expected average peers' punishment. This result seems to be driven by those bystanders that always decide not to punish dictators. I also found evidence that subjects in TREATED are more likely to change punishment decision across periods. Moreover, in CONTROL the amount of punishment in the two periods remains constant, while in TREATED it decreases. The mean punishment difference across periods is statistically higher in CONTROL than in TREATED if I consider situations where dictators take positive amounts from receivers' endowment. Finally, a large absolute difference between a bystander's beliefs regarding peers' punishment and actual peers' average punishment increases her likelihood to modify punishment decisions across periods.

These results provide evidence against the Zero Social Influence hypothesis. Hence, I conclude that the Zero Social Influence hypothesis is not supported by the results of the experiment.

4.5.2. Differential Social Influence hypothesis

First, I want to verify if bystanders that engage in less punishment in the first period are also less responsive to social influence. Third-parties receive relevant social information in the NORMATIVE and INFORMATIONAL treatments only, so I restrict the analysis to these treatments. For each of the seven dictators' decisions, I characterize bystanders that punish in the first period above the median as "high punishers"[71]. For each transfer level considered, the percentage of third-parties modifying punishment decision across periods among high punishers is almost double of that among the other punishers. If I exclude selfish bystanders I still have similar results.

I test the hypothesis implementing a logistic model. I estimate the probability of modifying punishment decision including the independent variable $Strat\_Punish$ that reports the level of punishment provided in the first period. Results are reported in Table 7. In any model specification, the coefficient associated with $Strat\_Punish$ is positive and significant at the 1% level. The coefficient of $Strat\_Punish$ suggests that, holding constant at their means the other controls, a bystander spending 1 additional token in first period punishment is 3% to 5% more likely to revise her punishment choice in the second period.

Hence, I conclude that this first set of results supports the Differential Social Influence hypothesis.

Second, I want to verify if bystanders that choose to punish in the first period a quantity different from their beliefs regarding average peers' punishment are less responsive to social influence compared to the other bystanders. I implement a logistic regression estimating the probability that a bystander modifies punishment decisions across periods. As independent variable, I introduce $Abs\_p0Belifs$, the absolute difference between a bystander's beliefs regarding average peers' punishment and her individual punishment in the

[71] Results are substantially the same if I choose the average punishment as a criterion for classification.

first period. I report results in Table 7.
The coefficient of $Abs\_p0Belifs$ is negative and statistically different from 0 in model specification 7, in which I include all the controls[72]. The estimations suggest that increasing the absolute difference $Abs\_p0Belifs$ by one unit decreases for a bystander the probability of modifying the punishment decision across periods by approximally 3%.
Therefore, I conclude that also this second set of results supports the Differential Social Influence hypothesis

- Conclusion relative to the Differential Social Influence hypothesis: *There is evidence that the more a bystander punishes in the first period, the more she is responsive to the social information received. I also find evidence that the larger the absolute difference between a bystander's beliefs regarding peers' average punishment and her first period punishment is, the less likely it is that she modifies punishment decisions across periods.*
  Therefore, I conclude that the results of my experiment support the Differential Social Influence hypothesis.

4.5.3. Equivalence of Normative and Informational Influence hypothesis
I conclude this section reporting results on the difference between bystanders exposed to both normative and informational social influence and those exposed only to the latter. From the summary statistics reported in Table 5, I could see that in the first period third-parties in INFORMATIONAL punish on average 4.1 tokens versus 3.3 of those in NORMATIVE. This difference is not statistically significant (p-value 26%). In both treatments, on average bystanders reduce punishment between the first and the second period: NORMATIVE of 0.28 tokens and INFORMATIONAL of 0.33. Also this difference is not statistically significant.

[72]In model 8 I exclude from the analysis selfish bystanders and the coefficient becomes not statistically significant.

I test the hypothesis that third-parties in NORMATIVE are more likely to revise their second period punishment decisions. I create the dummy variable $NORMATIVE$ equal to 1 for third-parties in the normative treatment and I implement a logistic regression. The dependent variable $DummyP1p0$ is equal to 1 when punishment is modified across periods. Results are reported in Table 9.

From the coefficient of $NORMATIVE$ in model 1 to 4 I could see that, on average, there is no statistical difference between treatments in the likelihood of modifying punishment decision. When instead I disentangle the effect of individual determinants of the probability to modify punishment decision across periods it is possible to find differences between treatments. First, consider the tests I did for the Differential Social Influence hypothesis. Models 7 and 8 of Table 7 suggest that increasing the absolute difference across punishment in the first period and individual beliefs regarding the average punishment in the session (the variable $Abs\_p0Belifs$) by one unit decreases the likelihood to modify punishment between period by approximately 3%. However, the result is only weakly significant. Nevertheless, the estimation could be affected by the fact that in the models of table 7 I constrained the slope of $Abs_p0Belifs$ to be the same for NORMATIVE and INFORMATIONAL. Therefore, in model 3 and 4 of Table 9 I introduce the interaction term $NorAbs\_p0Belifs$, that isolates the effects of the absolute difference between punishment in the first period and beliefs about peers' average punishment for bystanders in the NORMATIVE treatment. The coefficient is positive and significant at the 1% level for both model specifications, and the coefficient of $Abs_p0Belifs$ becomes negative and significant at the1% level. Interpreting the coefficients, I can see that for third-parties in NORMATIVE $Abs\_p0Belifs$ has no effect on the probability of modifying punishment across periods. Instead, for bystanders in INFORMATIONAL an increase of one unit in $Abs\_p0Belifs$ diminishes the probability of modifying punishment across periods by roughly 8%. Therefore, the results contrast the Differential

Table 9: Probability Modify Punishment Across Periods: Treated Groups

	(1)	(2)	(3)	(4)
NORMATIVE	0.064	0.075	0.034	0.154
	(0.11)	(0.09)	(0.17)	(0.15)
Strat_Punish			0.054***	0.039**
			(0.01)	(0.02)
Abs_p0Belifs			-0.085***	-0.091***
			(0.02)	(0.03)
NorStratPunish			0.016	0.014
			(0.02)	(0.02)
NorAbs_p0Belifs			0.103***	0.116***
			(0.03)	(0.04)
Abs_Signalp0			-0.003	0.001
			(0.02)	(0.02)
NorAbs_Signalp0			0.007	-0.003
			(0.02)	(0.02)
Abs_BelAvgPun			0.094***	0.094***
			(0.02)	(0.03)
NormAbs_BelAvgPun			-0.104***	-0.117***
			(0.03)	(0.04)
DummySignalp0			0.064	0.152
			(0.10)	(0.11)
NorDummySignalp0			0.095	-0.038
			(0.14)	(0.17)
Other contr	Y	Y	Y	Y
N	609	420	609	420
pseudo R^2	0.092	0.092	0.316	0.221

Notes: Logistic regression: dep. var. $DummyP1p0$, marginal effect at means, SE clustered by subject. Significance levels: * p<0.10, ** p<0.05, *** p<0.01. Other Controls include: male age degree worker social arts field_other risk logic impulsivity Instruction DictatorTake.

Social Influence hypothesis for bystanders in the NORMATIVE treatment, while the hypothesis finds support for subjects in the INFORMATIONAL treatment.

As a possible explanation for this difference across treatments, I conjecture

that for bystanders in NORMATIVE there are additional incentives to revise their punishment decisions compared to bystanders in INFORMATIONAL. In fact, in NORMATIVE bystanders are told that their punishment choices of the second period will be revealed to other participants and that these peers will express their judgements. Therefore, it seems that the threat of revealing individual choices to other participants triggers the decision to modify first period punishment.

Furthermore, consider the absolute difference between a bystander's beliefs regarding peers' average punishment and the actual peers' average punishment. Models 7 and 8 of Table 7 indicate that increasing the coefficient of the variable $Abs\_BeliefAvgPunish$ by one unit increases for a bystander the probability to modify punishment decision across periods by 3%. However, this result comes from models where I constrained the coefficient of $Abs\_BeliefAvgPunish$ to be the same in NORMATIVE and INFORMATIONAL.

I verify if the coefficient is the same in both treatments estimating the effect of $Abs\_BeliefAvgPunish$ for the two groups separately. I do so interacting the variable $Abs_BeliefAvgPunish$ with the dummy NORMATIVE and so creating the variable $NormAbs\_BeliefAvgPunish$. From the results of models 3 and 4 in Table 9, we can see that the coefficient of $NormAbs\_BeliefAvgPunish$ is negative and statistical significant at the 1% level. On the other hand, the coefficient of $Abs\_BeliefAvgPunish$ in the unconstrained model becomes positive and significant at the 1% level, while it was only weakly statistically significant in the constrained model. Specifically, for subjects in the INFORMATIONAL treatment an increase of one unit in $Abs\_BeliefAvgPunish$ raises the probability of modifying punishment across periods by approximately 9%.

In order to further investigate this result, I check how bystanders in the two treatments modify their punishment choices across periods conditional to the sign of the difference between beliefs regarding peers' average punish-

ment and actual peers' average punishment. Table 8 reports these summary statistics. Bystanders in both treatments reduce punishment in the second period when actual average punishment is lower than expected. However, bystanders in INFORMATIONAL on average reduce punishment by more than 1 token, while those in NORMATIVE by less than 0.5. Conversely, when actual average peers' punishment exceeds a bystander's expectations, in INFORMATIONAL third-parties increase punishment by 0.4 tokens on average, while bystanders NORMATIVE do not modify punishment decisions.

It is possible that the lower variability registered in NORMATIVE derives from the fact that individual choices are observable by peers. I conjecture that in NORMATIVE bystanders refrain from modifying punishment decisions, in particular from reducing punishment, because of the disutility of being eventually judged and targeted with the "sad" emoticon by peers.

Finally, I also test if the slope of the variable $Strat\_Punish$ differs between NORMATIVE and INFORMATIONAL. I created the variables $NorStrat\_Punish$, isolating the effect of $Strat\_Punish$ for bystanders in the NORMATIVE treatment. As expected, results for these unconstrained models reported in Table 9 show that there is no statistical difference between treatments in the data.

- Conclusion relative to Equivalence of Normative and Informational Influence hypothesis: *I find mixed evidence regarding my hypothesis. On the one hand, at an aggregate level the likelihood to modify punishment choices is the same in NORMATIVE and INFORMATIONAL. However, disentangling the determinants that push bystanders to modify punishment choices across periods, I find differences between the two treatments.*

  Therefore, I conclude that the empirical evidence is mixed and the hypothesis is not fully supported by the data.

4.6. Conclusions of Chapter 4

Human organizations need mechanisms to enforce rules and regulations upon which they are founded. On the one hand, societies have developed a centralized apparatus of enforcement for this purpose. However this centralized systems coexist with a decentralized practice of punishment carried out by members of the society itself. Understanding the nature and characteristics of decentralized punishment might help legal scholars and policymakers to design effective policies in a variety of situations. Therefore, which are the major drivers of decentralized third-party punishment is an important question for social scientists.

In this chapter I examine through a laboratory experiment the effect of one of these drivers, social influence, on the punishment decisions of third parties. Scholars in psychology, law and economics underline the relevance of third party punishment for the cohesion of human societies (Fehr and Fischbacher, 2004b; Marlowe et al., 2008) and the importance of social influence in various fields of application (Bernheim, 1994; Turner, 1991; Kahan, 1997; Becker, 1991). However, this paper is the first work that investigates empirically social influence effects within the framework of third party punishment.

In a modified dictator game, I elicit the punishment choices of third parties before and after having exposed them to information regarding the punishment behavior of their peers. I compare those choices with decisions made by bystanders not exposed to social relevant information. The main finding of this chapter is that social influence is a major driver of bystanders' decision to engage in third-party punishment. Results of the experiment show that third-parties receiving information about peers' punishment revise their punishment choices more often and on average reduce punishment across periods compared to bystanders exposed to social irrelevant information. This last effect seems to be driven by the fact that bystanders' beliefs regarding peers' average punishment are higher than the actual punishment peers exert. Indeed, consistently with the model predictions, the empirical analysis shows

that the larger the absolute difference between a bystander's beliefs about peers' average punishment and peers' actual punishment is, the more likely the bystander is to revise her initial decisions.

I also disentangled the effect of two possible channels of social influence. Results suggest that some third-parties are only responsive to the discomfort of disagreeing with the majority, that is at the base of normative social influence and their punishment choices are not influenced by the "need to be right" on which informational social influence is based. Distinguishing between these two channels of social influence is of primary importance for social analysts, since previous studies document that informational social influence causes a permanent change in behavior (see for example Newcomb et al., 1967). On the other hand, normative social influence is more ephemeral and leads to modifications of behavior that are subject to specific circumstances[73] (Deutsch and Gerard, 1955; Cohen and Golden, 1972; Burnkrant and Cousineau, 1975).

These findings have two major implications. On the one hand, they stress the importance in our societies of citizens' perception about peers' behavior. This is expecially important in situations where beliefs of the general population systematically overestimate the frequency of socially undesirable behaviors, like frequently happens for perceived crime, benefit frauds or the percentage of non-voters[74]. In these situations, policymakers might often achieve welfare-improving results by means of ad-hoc communication strategies that could outperform alternative and often more costly policies (see for example Casal and Mittone, 2014, where the authors discuss an application

[73]Nevertheless, scholars proposed models of endogenous preferences, arguing that even individuals initially adopting compliant behaviors by means of normative social influence may endogenously modify their preferences (Akerlof, 1989; Klick and Parisi, 2008).

[74]For example, the Royal Statistical Society reports that 58% of the UK population estimates that crime is rising, while data show how crime rate in the country is 19% lower than the previous year and 53% lower than 1995. For discussion of other examples and additional details see http://www.kcl.ac.uk/newsevents/news/newsrecords/2013/07-July/Perceptions-are-not-reality-the-top-10-I-get-wrong.aspx.

of social stigma to tax evasion).

On the other hand, even when population beliefs are not biased, the possibility of resorting to social influence as a subsidiary tool for achieving compliance has been advanced by scholars in an array of situations of economic importance (Ela, 2008; Posner, 2000; Cooter, 1998; Zasu, 2007). As a society, we invest a considerable amount of resources with the objective of shaping individual beliefs and direct them toward social desirable outcomes. Policymakers might want to encourage, by means of a social influence approach, third party interventions in situations where the lack of resources prevent a centralized authority to perform effective interventions. This is the case for example of the recent campaign aiming at prevention of social offenses *"Bringing in the Bystander"* promoted in the UK by the National Sexual Violence Resource Center. The campaign aims at reducing social offenses employing a marketing campaign that explicitly encourages third parties intervention[75].

I agree with Mathew and Boyd (2011) that third-party punishment represents "the cement of human societies". In this chapter I argue for the first time about the possibility for policymakers to take advantage of social influence effects in promoting third-party punishment, reporting evidence from a laboratory experiment that social influence significatively affects bystanders' interventions. Given the importance and wide possibilities of application in the societal framework, I hope that future researches further investigate the conection between social influence and third party punishment, in particular verifying the robustness of my findings in a field setting and the persistence of the effects in a longer term horizon.

[75]"Using a bystander intervention approach combined with a research component, this program assumes that everyone has a role to play in prevention [...] The Know Your Power campaign is the social marketing component of Bringing in the Bystander".

5. Conclusions

Government policy interventions might greatly improve the welfare of a society. However, they could also result ineffective, wasting taxpayers money without producing desired outcomes or be unreasonably invasive, reducing people's freedom of choice. Therefore, creating policies that on the one hand are cheap to implement and on the other hand identifying those contexts where they will prove effective are two major challenges for the policy analyst.

This book aims at discussing and contributing to behavioral public policymaking, a social policy movement that, according to many scholars, satisfies both the requirements. Behavioral policies address systematic and predictable violations of rationality to steer agents' behavior in directions that are welfare-improving. The nonstandard behavioral regularities that the behavioral policy analysts identify and exploit are natural characteristics of human decision-making process. Therefore, policy interventions proposed are relatively easy and not expensive to implement, since they take advantage of already established patterns of behavior to achieve the policymaker's goals. Moreover, a specific approach within the behavioral public policymaking movment, the so called "Libertarian Paternalism", is respectful of the individual freedom of choice if compared to the classical forms of paternalistic interventions. In fact, a key feature of libertarian paternalistic policies is that they neither mandate to individuals any specific behavior nor they increase the monetary costs of selecting certain outcomes. Therefore, libertarian paternalism does not restrict agents' choice set and minimize the risk to reduce individuals' welfare by constraining behaviors.

In the introductory chapter, I focus in particular on the fundamental methodological problem of behavioral public policymaking, which is finding a suitable welfare criterion. I stress how the behavioral analyst cannot perform welfare analysis simply relying in the revealed preference principle used in neoclassical economics. I also point out how scholars did not reach a consen-

sus yet and how the debate in this area is evolving.

Moreover, in the last section of the introduction I argue that behavioral scientists can and should exert more effort in order to influence policymaking. I stress that scholars should balance the trade-off between scientific exactness and policymaker's need of clear and precise policy suggestions. I also point out that behavioral scientists need to engage in more field experimentation and that they should be able to reconcile theoretical and empirical works in an unitary framework of analysis.

However, the main goal of this book is not to discuss methodological issues but instead to contribute to the development of the behavioral public policymaking movement by proposing methodological advances and original contributions. In the next three sections I summarize my findings relative to chapter 2, 3 and 4. In section 5.4 I then conclude underlying the academic relevance and the policy implications of my results.

5.1. Chapter 2: Summary of Findings

In this chapter I focused on a critical methodological problem faced by policy analysts, that is the aggregation of individuals' well-being in a unitarian measure of social welfare. The anaytical tool that is used for aggregating individuals' well-being in an unitarian measure is called the social welfare function. The social analyst that has to construct a social welfare function follows a two-step procedure. As a first step, he has to make interpersonal comparisons of utility and subsequently he aggregates the measures of individual utilities. I discussed how the choices made by the social analyst in both these steps reflect different normative value judgements. I then summarized the different positions embraced by scholars that discussing which is the most ethically and philosophically appropriate set of value judgements for conducting welfare analyses. I underlined how the debate in this area is still open and how the different positions embraced by scholars reflect the heterogeneity in preferences for efficiency versus redistribution.

I then proposed my contribution to the debate. I argued that, while different methods for aggregating individuals' well-being reflect different preferences for redistributions, nevertheless it is possible to identify quantitative relationships between the social welfare functions that the social analysts use. There are no contributions neither in law nor in economics attempting to identify these relationships. Therefore, I analyze the quantitative relationship between the forms of social welfare function most commonly used by social analysts.

As a first result, I formally showed that in general different social welfare functions do not necessarily produce the same policy evaluation results. This result might be intuitive for economists, however I am not aware of any contribution proposing a formal proof of it. Therefore, I decided to close this gap. Moreover, I showed that a subset of social welfare functions represents an exception to my previous general results. In fact, I identified what are the social welfare functional forms that in a policy evaluation result always produce the same result.

After this, in the core of this chapter, I derive a quantitative conditions necessary to generalize the policy evaluation results obtained implementing a specific combination of individual utility and aggregation method to the other social welfare functional forms considered. In this part, I also showed that it is possible to derive general results if we impose some restrictive conditions on the transfer of resources implied by the policy under consideration. I performed this analysis and I derived the quantitative results both in a two-agent scenario and in a two-interest group situation. My results show how the quantitative relationship between different social welfare functions vary as a function of the number of agents composing the interest groups.

5.2. Chapter 3: Summary of Findings

In the rest of the thesis I proposed two original contributions that discussed the possibility to implement two libertarian paternalistic policies. In chapter 3 I discuss a zero-cost policy intervention based on stochastic rewards

that aims at combatting indirect tax evasion. In many countries, indirect tax evasion represents an endemic problem and creates difficulties to the sustainability of public finances. One major way to evade indirect taxes for business owners consists in not releasing to customers invoices that register business transaction The policy object of this chapters suggests the introduction by the government of a lottery that links the possibility to win a stochastic prize to the possession of an invoice. Therefore, customers are incentivized to request the emission invoices to business owners.

Theoretical predictions derived from models of standard decision-making state that this reward mechanism cannot be effective. In fact, according to these models predictions, the reward that the government should offer to customers in order to incentivize them to enforce invoices emission has to be necessarily higher than the aditional tax revenue collected as a consequence of the lottery introduction, resulting in a loss for the government. Despite these predictions, this policy has been applied in few countries. Empirical estimations of the effect of the policy introduction show that it is effective in reducing indirect tax evasion and that the increase in tax collected more than compensates the cost of the prize paid by the government. However, it has not been explained why this policy proved to be successful and the empirical evidence remains puzzling.

The objective of chapter 3 is to propose an empirical model that explains the empirical evidence and that endows policymakers interested in applying the lottery policy with a theoretical framework to predict the policy effects. I started discussing the empirical evidence showing the policy success in the countries where it has been applied. I then showed that models of expected utility, the benchmark for formal analysis of individual decision-making, fail to explain these results. I then proposeed my original contribution. I presented a model based on a theory of non-expected utility that incorporates people's behavioral tendency to overweight small probabilities in risky choices. I formally showed that the predictions generated by my

model are consistent with the empirical evidence registered. I then proposed a calibration exercise and I derived the conditions necessary to register an effective implementation of the lottery policy.

I concluded the chapter discussing side-effects connected with the introduction of the policy. On the one hand, there is the risk that the introduction of the lottery policy could crowd-out the voluntary enforcement of invoice emission carried out by ethically motivated citizens. Hence, the crowding-out effect could potentially offset benefits deriving from the policy. On the other hand, there are some positive side-effects connected with the lottery introduction. First, even if the increase in tax revenue collected does not compensate the prize paid out by the government and so the policy must be abandoned, data collected during the time period when the policy is in place might help the screening process of the authority sanctioning indirect tax evasion. In fact, business whose owners were evading taxes would register a spike in reported revenue when the policy is in place, since some customers request the invoices emission. Therefore, these businesses could be selectively targeted and monitored by the tax authority. This would increase the effectiveness of the monitoring and sanctioning process and it would reduce the amount of taxes evaded. Moreover, I argued that the lottery policy, even if introduced for a limited period of time, might change people's norm of behavior with respect to enforcing invoices emission. If citizens become used to asking for an invoice because of the opportunity to win the lottery prize, they will possibly continue to do it even absent the prospect of a prize. In fact, inertia and social norms of behavior once in place tend to persist in the population. Therefore, the lottery policy might be implemented as a temporary intervention that incentivizes people reticent to change behavior to modify the status quo in favor of a welfare-improving alternative.

5.3. Chapter 4: Summary of Findings

In chapter 4 I investigated the possibility of creating a policy that exploits social influence effects. I focused on the effects of social influence on bystanders'

likelihood to engage in costly punishment or, in the language of economics, "decentralized third-party punishment". Decentralized third-party punishment is a form of altruistic intervention carried out by a private agent that incurs material costs for sanctioning the behavior of a wrongdoer, even if the punisher is not directly affected by the wrongdoer's action. I discussed the importance of third-party punishment for the existence of human organizations and I underlined that some scholars even consider it "the cement of societies". While social sciences have extensively investigated aspects of both social influence and third-party punishment, the contribution offered in this chapter is the first that formally studies their interconnection.

My goal in this chapter was twofold. On the one hand, I proposed a model of decision-making that takes into account social influence effects. This model allows to derive sharp theoretical predictions that could be tested empirically. On the other hand, I had to empirically test my model predictions and showed that social influence causally affects punishment behavior of the third-parties. Moreover, I also wanted to identify the channels of transmission of social influence to individuals.

To achieve the first goal, I proposed a theoretical model of social influence. According to my model, social influence agents exposed to social influence would modify their individual behavior and choose actions that deviate from the theoretical predictions of models of standard decision-making. I then proceeded testing my model predictions. However, isolate social influence effects is a challenging task and estimations of this effects performed with field data suffer serious identification problems. Therefore, I exploited the possibilities offered by the controlled environment of a laboratory experiment for ruling out confounding factors and self-selection problems. I designed and ran a laboratory experiment following the methodology of experimental economics.

Results of my experiment are consistent with the predictions of my model. I showed that social influence is a major driver of decentralized third-party

punishment. Third-parties exposed to social influence are significatively more likely to modify their punishment decisions compared to bystanders exposed to irrelevant social information. In particular, I showed that the more the initial beliefs of a third-party regarding their peers' punishment choices is incorrect, the more she is likely to change her punishment decision taken in isolation when exposed to information regarding actual peers' punishment. Moreover, I identified the two channels through which social influence operates. Contributions in psychology distinguish between normative social influence, that corresponds to "the need to be liked" by peers and informational social influence, that fulfill an agent's "need to be right". I show that some agents respond to the former type of social influence but not to the latter. Disentangling the effects of these two channels of social influence is relevant for policy purposes. In fact, researches show that the effect on behavior of informational social influence are more persistent than those of normative social influence.

5.4. Academic Relevance and Policy Implications

In this book I contributed to the discussion on the rising movement of behavioral public policymaking. While the book contains a general introduction that extensively discusses the ideas behind this social policy movement, its critiques and fields for future research, nevertheless my main goal was to produce original research within the behavioral policy framework. I tried to do so proposing an innovative methodological contribution and discussing two new behavioral policies. The results of my research has potential implications for scholars as well as for policymakers and decisionmakers.

In chapter 2, I derived the quantitative relationship existing between the most common forms of social welfare functions used for conducting social policy analysis. This chapter aims to further enrich the discussion regarding the choice of the most appropriate social welfare function in policy analysis, proposing an original methodological advance. The results derived in this chapter are relevant for scholars and researchers that employ formal social

welfare analysis. In fact, the quantitative relationships derived in this chapter allow to verify to what degree the predictions derived by the social analyst's model are sensitive to the set of value judgements implied by the social welfare function chosen. Moreover, results of this chapter are important for decisionmakers that have to make the final choice regarding the possibility to implement a policy. In fact, the quantitative relationships that I derived allow to identify how general and respectful of the different ethical positions in the population the result obtained by a policy analysts are. Indeed, we have seen that any policy analysis necessarily reflects the value judgements chosen by the policy analyst. Finally, I underline how the applicability of the results derived in this chapter is not limited to the area of behavioral public policy. Indeed, they are relevant also for social policy evaluations grounded in traditional welfarist economics and also for extra-welfarist analyses of social policy.

In chapter 3, I discussed a complementary policy to the traditional deterrence approach to VAT and RST evasion recently adopted by some Asian and Latin-American countries. This policy incentivizes customers to enforce invoices emission from sellers by linking the possession of an invoice to the possibility to win a stochastic prize. The primary objective of this chapter was to provide a theoretical model that explains why, contrary to standard predictions, the lottery policy results successful in increasing VAT and RST net revenue. I proposed a model based on Tversky and Kahneman's Cumulative Prospect Theory (1992) that is able to explain the puzzling empirical evidence. Given the model specification and calibration of parameters, I then introduced a test for verifying the applicability of the lottery in the specific environment and population of interest. The implication of my results are relevant for policymakers interested in applying the lottery policy. In fact, I argue that risk preferences and social norms of behavior are key elements to be taken into account by policymakers in forecasting the policy effectiveness. My contribution consisted in having provided a theoretical framework

that enables policymakers to generate sharp predictions based on empirical measures. Therefore, the theoretical framework and the test I propose represent useful ex-ante indicators of the expected success of the lottery policy in increasing the level of indirect tax compliance.

Finally, in chapter 4 I showed that social influence is a major determinant of third-party punishment and I argued that the effectiveness of policies based on bystanders' intervention could be enhanced by the effects of social influence. The implications of the results derived in this chapter are twofold. First, they suggest policymakers to use a social influence approach to encourage third party interventions in situations where the lack of resources prevent a centralized authority to perform effective interventions. Campaigns aimed at preventing bullyism or protection of victims of social offenses are possible fields of application for these plocies. More generally, my contribution suggests the possibility to promote libertarian paternalistic policies based on social influence in situations where agents' perception of the frequency of an event is wrong. In fact, it is often the case that agents systematically underestimate the frequency of welfare-improving actions undertaken by their peers. In these situations, results of my work show that the government could increase the frequency of the welfare-improving behavior promoting policies that convey correct information to the population. These policy interventions are not expensive, easy to implement and would achieve long-lasting effects, since affect agents' behavior through the channel of informational social influence.

I am confident that the rising field of behavioral public policy have the pontential to further expand and dramatically improve people's life. However, a great amount of work has still to be done in order to meet this objective. I hope that behavioral scientists and policymakers will increasingly devote their energies to research in this area. I also hope that future quantitative researches on the choice of the social welfare function will be conducted. In particular, it would be interesting to extend the analysis presented in chap-

ter 2 to other forms of social welfare functions and to verify if it is possible to generalize additional results. I also hope that the model presented in chapter 3 will direct the attention of researchers to the lottery ticket policy. In particular, since some countries are planning to implement this policy in the immediate future, it would be interesting to verify my model predictions empirically. Finally, I hope that future researches further investigate the conenction between social influence and third party punishment that I presented in chapter 4. In particular, I hope scholars will be able to verify the robustness of my findings in a field setting and the persistence of the effects in a longer term horizon.

Appendix A.

Proof of Proposition 1

I show my proposition using a proof by contradiction, that is a is a particular kind, often use in mathematics, of the more general form of argument known as "reductio ad absurdum". Without any loss of generality, I define SW^1 as the SWF with the smaller ratio: $\frac{\frac{\partial SW^1}{\partial w_j}}{\frac{\partial SW^1}{\partial w_i}} < \frac{\frac{\partial SW^2}{\partial w_j}}{\frac{\partial SW^2}{\partial w_i}}$
Define $W = (w_1, w_2, ..., w_n)$ and $W^\epsilon = (w_1, ..., w_i + k\epsilon, ..., w_j - \epsilon, ..., w_n)$ where k and ϵ are positive real numbers. We want to show that there exist values of k and ϵ such that $SW^1(W^\epsilon) > SW^1(W)\, and\, SW^2(W^\epsilon) < SW^2(W)$. To show our result, first notice that for small enough ϵ:

$SW^1(W^\epsilon) \approx SW^1(W) + \frac{\partial SW^1}{\partial w_j}(-\epsilon) + \frac{\partial SW^1}{\partial w_i}k\epsilon$ and $SW^2(W^\epsilon) \approx SW^2(W) + \frac{\partial SW^2}{\partial w_j}(-\epsilon) + \frac{\partial SW^2}{\partial w_i}k\epsilon$

If we can show that there exists a value of k such that:

(I) $SW^1(W) + \frac{\partial SW^1}{\partial w_j}(-\epsilon) + \frac{\partial SW^1}{\partial w_i}k\epsilon > SW^1(W)$
and

(II) $SW^2(W) + \frac{\partial SW^2}{\partial w_j}(-\epsilon) + \frac{\partial SW^2}{\partial w_i}k\epsilon < SW^2(W)$

then, by continuity of the SWF, we can claim that for small enough ϵ: $SW^1(W^\epsilon) > SW^1(W)$ and $SW^2(W^\epsilon) < SW^2(W)$; which proves our proposition.

From (I) and (II), it follows that:

$$k > \frac{\frac{\partial SW^1}{\partial w_j}}{\frac{\partial SW^1}{\partial w_i}}$$

$$k < \frac{\frac{\partial SW^2}{\partial w_j}}{\frac{\partial SW^2}{\partial w_i}}$$

However, by assumption, $\frac{\frac{\partial SW^1}{\partial w_j}}{\frac{\partial SW^1}{\partial w_i}} < \frac{\frac{\partial SW^2}{\partial w_j}}{\frac{\partial SW^2}{\partial w_i}}$. Hence, there must be a range of values of k for which conditions (I) and (II) hold, thus proving our proposition.

Proof of Corollary 1

From proposition 1, we just need to show that for some i and j, $\frac{\frac{\partial SW^1}{\partial w_j}}{\frac{\partial SW^1}{\partial w_i}}$ is different for each criteria considered.

Now let any utility function be the maximand of each SWF. Functions' partial derivatives are:

- Bentham criteria: $\frac{\partial SWB(W)}{\partial w_j} = u'(w_j)$
- Nash criteria: $\frac{\partial SWN(W)}{\partial w_j} = u'(w_j)\Pi_{i \neq j}u(w_i)$
- Rawls criteria: $\frac{\partial SWR(W)}{\partial w_j} = \begin{cases} 1, \ if \ w_j < w_i \ \forall i \\ 0, \ if \ w_j \geq w_i \ for \ some \ i \end{cases}$

Therefore, the ratio of derivatives is:

- $\frac{\frac{\partial SWB}{\partial w_j}}{\frac{\partial SWB}{\partial w_i}} = \frac{u'(w_j)}{u'(w_i)}$
- $\frac{\frac{\partial SWN}{\partial w_j}}{\frac{\partial SWN}{\partial w_i}} = \frac{u'(w_j)}{u'(w_i)} \frac{u(w_i)}{u(w_j)}$
- $\frac{\frac{\partial SWR}{\partial w_j}}{\frac{\partial SWR}{\partial w_i}}$ = either 0 or not existent.

  The case where wealth is defined as the maximand can be simply considered a subcase where $u(w_j) = w_j$ and $u'(w_j) = 1$

Given that the three derivative results are different, the three welfare criteria are not equivalent when the same maximand is chosen.

Proof of Proposition 2

Since a SWF ranks states in an ordinal sense, any monotonic transformation does not affect the ranking order. Therefore, if I show that by undertaking monotonic transformations on a SWF among those described in proposition 2 I can achieve any of the others' SWF form, I show that all the three specifications yield the same rank.

As a starting point, consider a Nash-wealth SWF:

$\text{SWN}^w(W) = \Pi_{i=1}^{N} w_i$

by raising to the power of α we obtain:

$= \Pi_{i=1}^{N} w_i^{\alpha_1} = SWN^{pol}(W)$

this proves the equivalence with SWF Nash Polynomial.
I continue by taking the logarithm and dividing it by α:

$= \sum_{i=1}^{N} lnw_i = SWB^{log}(W)$

this proves the equivalence with SWF Bentham Logarithmic.

Since all of the transformations above are monotonic, this implies that all three specifications rank alternative states of the world in the same way, and are therefore equivalent.

Table A.10: **Minimal Level of K required to guarantee Welfare Improvement - Homogeneous Groups**

Utility Func. Form	Bentham	Nash	Rawls
Wealth	1	$\frac{w_A}{w_B-\epsilon}$	$\emptyset$
Polynomial	$(1/\epsilon)\{[w_A^\alpha + w_B^\alpha - (w_B-\epsilon)^\alpha]^{1/\alpha} - w_1\}$	$\frac{w_A}{w_B-\epsilon}$	$\emptyset$
Logarithmic	$\frac{w_A}{w_B-\epsilon}$	$(1/\epsilon)[e^{\frac{lnw_A lnw_B}{lnw_B-\epsilon}} - w_A]$	$\emptyset$
Exponential	$(1/\epsilon)\{-(1/\alpha)ln[e^{-\alpha w_A} + e^{-\alpha w_B} - e^{-\alpha(w_B-\epsilon)}] - w_A\}$	$\frac{1}{\epsilon}\left\{\left(\frac{-1}{\alpha}\right) ln\left[1 - \left[\frac{(1-e^{-\alpha w_A})(1-e^{-\alpha w_B})}{1-e^{-\alpha(w_B-\epsilon)}}\right]\right] - w_A\right\}$	$\emptyset$

Table A.11: **Minimal Level of K required to guarantee Welfare Improvement - Non-homogeneous Groups**

Utility Func.	Bentham	Nash	Rawls
Wealth	$\frac{N-j}{j}$	$(1/\epsilon)[(\frac{w_A^j w_B^{N-j}}{(w_B-\epsilon)^{N-j}})^{(1/j)} - w_A]$	$\emptyset$
Polyn	$(1/\epsilon)\{[\frac{jw_A^\alpha+(N-j)w_B^\alpha-(N-j)(w_B-\epsilon)^\alpha}{j}]^{1/\alpha} - w_A\}$	$(1/\epsilon)[(\frac{w_A^j w_B^{N-j}}{(w_B-\epsilon)^{N-j}})^{(1/j)} - w_A]$	$\emptyset$
Log	$(1/\epsilon)[(\frac{w_A^j w_B^{N-j}}{(w_B-\epsilon)^{N-j}})^{(1/j)} - w_A]$	$(1/\epsilon)\{e^{[\frac{(lnw_A)^j (lnw_B)^{N-j}}{(lnw_B-\epsilon)^{N-j}}]^{1/j}} - w_A\}$	$\emptyset$
Exp	$(1/\epsilon)\{-(1/\alpha)ln[\frac{je^{-\alpha w_A}+(N-j)e^{-\alpha w_B}-(N-j)e^{-\alpha(w_B-\epsilon)}}{j}] - w_1\}$	$(1/\epsilon)\{(\frac{-1}{\alpha})ln[1-(\frac{(1-e^{-\alpha w_A})^j(1-e^{-\alpha w_B})^{N-j}}{(1-e^{-\alpha(w_B-\epsilon)})^{N-j}})^{1/j}] - w_A\}$	$\emptyset$

Proofs Tables A.10

1. Bentham Case:

   - Wealth: $u(w_i) = w_i$

     Proof.

$$SW(W') \geq SW(W) \tag{A.1}$$
$$\iff (w_1 + k\epsilon) + (w_2 - \epsilon) \geq w_1 + w_2 \tag{A.2}$$
$$\iff k \geq 1 \tag{A.3}$$

     □

   - Polynomial: $u(w_i) = w_i^{\alpha}$

     Proof.

$$SW(W') \geq SW(W) \tag{A.4}$$
$$\iff (w_1 + k\epsilon)^{\alpha} + (w_2 - \epsilon)^{\alpha} \geq w_1^{\alpha} + w_2^{\alpha} \tag{A.5}$$
$$\iff k \geq \frac{1}{\epsilon}\{[w_1^{\alpha} + w_2^{\alpha} - (w_2 - \epsilon)]^{1/\alpha} - w_1\} \tag{A.6}$$

     □

   - Logarithmic: $u(w_i) = ln(w_i)$

Proof.

$$SW(W') \geq SW(W) \tag{A.7}$$
$$\iff ln(w_1 + k\epsilon) + ln(w_2 - \epsilon) \geq ln(w_1) + ln(w_2) \tag{A.8}$$
$$\iff ln[(w_1 + k\epsilon)(w_2 - \epsilon)] \geq ln(w_1 w_2) \tag{A.9}$$
$$\iff (w_1 + k\epsilon)(w_2 - \epsilon) \geq w_1 w_2 \tag{A.10}$$
$$\iff k \geq \frac{w_1}{w_2 - \epsilon} \tag{A.11}$$

□

- Exponential: $u(w_i) = 1 - e^{-\alpha w_i}$

  Proof.

$$SW(W') \geq SW(W) \tag{A.12}$$
$$\iff (1 - e^{-\alpha(w_1 + k\epsilon)}) + (1 - e^{-\alpha(w_2 - k)}) \geq (1 - e^{-\alpha(w_1)}) + (1 - e^{-\alpha(w_2)}) \tag{A.13}$$
$$\iff -e^{-\alpha(w_1 + k\epsilon)} - e^{-\alpha(w_2 - k)} \geq -e^{-\alpha(w_1)} - e^{-\alpha(w_2)} \tag{A.14}$$

  then dividing by $-e^{-\alpha w_1}$:

$$\tag{A.15}$$
$$\iff e^{-\alpha k\epsilon} \leq 1 + e^{-\alpha(w_2 - w_1)}(1 - e^{\alpha\epsilon}) \tag{A.16}$$
$$\iff k \geq (\frac{-1}{\alpha\epsilon}) ln[1 + e^{-\alpha(w_2 - w_1)}(1 - e^{-\alpha\epsilon})] \tag{A.17}$$

□

2. Nash Case:

- Wealth: $u(w_i) = w_i$

  Proof.

$$SW(W') \geq SW(W) \tag{A.18}$$
$$\iff (w_1 + k\epsilon)(w_2 - \epsilon) \geq w_1 w_2 \tag{A.19}$$
$$\iff k \geq \frac{w_1}{w_2 - \epsilon} \tag{A.20}$$

  □

- Polynomial: $u(w_i) = w_i^\alpha$

  Proof.

$$SW(W') \geq SW(W) \tag{A.21}$$
$$\iff (w_1 + k\epsilon)^\alpha (w_2 - \epsilon)^\alpha \geq w_1^\alpha w_2^\alpha \tag{A.22}$$
$$\iff (w_1 + k\epsilon)(w_2 - \epsilon) \geq w_1 w_2 \tag{A.23}$$
$$\iff k \geq \frac{w_1}{w_2 - \epsilon} \tag{A.24}$$

  □

- Logarithmic: $u(w_i) = ln(w_i)$

  Proof.

$$SW(W') \geq SW(W) \tag{A.25}$$
$$\iff ln(w_1 + k\epsilon) ln(w_2 - \epsilon) \geq ln(w_1) ln(w_2) \tag{A.26}$$
$$\iff ln(w_1 + k\epsilon) \geq \frac{ln(w_1) ln(w_2)}{ln(w_2 - \epsilon)} \tag{A.27}$$

raising both sides to the power e:

$$\iff w_1 + k\epsilon \geq exp[\frac{ln(w_1)ln(w_2)}{ln(w_2 - \epsilon)}] \tag{A.28}$$

$$\iff k \geq \frac{1}{\epsilon}\{exp[\frac{ln(w_1)ln(w_2)}{ln(w_2 - \epsilon)}] - w_1\} \tag{A.29}$$

□

- Exponential: $u(w_i) = 1 - e^{-\alpha w_i}$

  Proof.

$$SW(W') \geq SW(W) \tag{A.30}$$

$$\iff (1 - e^{-\alpha(w_1 + k\epsilon)})(1 - e^{-\alpha(w_2 - k)}) \geq (1 - e^{-\alpha w_1})(1 - e^{-\alpha w_2}) \tag{A.31}$$

$$\iff -e^{-\alpha(w_1 + k\epsilon)} + e^{-\alpha(w_1 + w_2 - k\epsilon - \epsilon)} \geq -e^{-\alpha w_2} - e^{-\alpha w_1} + e^{-\alpha w_1 - \alpha w_2} + e^{-\alpha(w_2 - \epsilon)} \tag{A.32}$$

$$\iff -e^{-\alpha k\epsilon}(e^{-\alpha w_1} + e^{-\alpha(w_1 + w_2 - \epsilon)}) \geq -e^{-\alpha w_2} - e^{-\alpha w_1} + e^{-\alpha w_1 - \alpha w_2} + e^{-\alpha(w_2 - \epsilon)} \tag{A.33}$$

$$\iff e^{-\alpha k\epsilon} \leq \frac{1 - e^{-\alpha w_2} + e^{-\alpha(w_2 - w_1)}}{1 + e^{-\alpha(w_2 - \epsilon)}} \tag{A.34}$$

$$\iff k \geq (\frac{-1}{\alpha\epsilon})ln[\frac{1 - e^{-\alpha w_2} + e^{-\alpha(w_2 - w_1)}}{1 + e^{-\alpha(w_2 - \epsilon)}}] \tag{A.35}$$

□

Proofs Table A.11

(a) Bentham Case:

- Wealth: $u(w_i) = w_i$

  Proof.

$$SW(W') \geq SW(W) \tag{A.36}$$

$$\iff \sum_{i=1}^{j}(w_i + k\epsilon) + \sum_{i=j+1}^{N}(w_i - \epsilon) \geq \sum_{i=1}^{N} w_i \tag{A.37}$$

$$\iff jk\epsilon - (N-j)\epsilon \geq 0 \tag{A.38}$$

$$\iff k \geq \frac{(N-j)}{j} \tag{A.39}$$

□

- Polynomial: $u(w_i) = w_i^{\alpha}$

Proof.

$$SW(W') \geq SW(W) \tag{A.40}$$

$$\iff \sum_{i=1}^{j}(w_i + k\epsilon)^\alpha + \sum_{i=j+1}^{N}(w_i - \epsilon)^\alpha \geq \sum_{i=1}^{N}(w_i)^\alpha \tag{A.41}$$

$$\iff j(w_A + k\epsilon)^\alpha + (N-j)(w_B - \epsilon)^\alpha \geq jw_A^\alpha + (N-j)w_B^\alpha \tag{A.42}$$

$$\iff k \geq \frac{1}{\epsilon}\left[\left(w_A^\alpha + \frac{(N-j)}{j}[w_B^\alpha - (w_B - \epsilon)^\alpha]\right)^{(1/\alpha)} - w_A\right] \tag{A.43}$$

$$\iff k \geq (1/\epsilon)\{[\frac{jw_A^\alpha + (N-j)w_B^\alpha - (N-j)(w_B - \epsilon)^\alpha}{j}]^{1/\alpha} - w_A\} \tag{A.44}$$

□

- Logarithmic: $u(w_i) = ln(w_i)$

Proof.

$$SW(W') \geq SW(W) \tag{A.45}$$

$$\iff \sum_{i=1}^{j} ln(w_i + k\epsilon) + \sum_{i=j+1}^{N} ln(w_i - \epsilon) \geq \sum_{i=1}^{N} ln(w_i) \tag{A.46}$$

$$\iff ln(w_A + k\epsilon)^j \geq ln(w_A^j w_B^{(N-j)}) - ln(w_B - \epsilon)^{(N-j)} \tag{A.47}$$

$$\iff w_A + k\epsilon \geq \left(\frac{w_A^j w_B^{(N-j)}}{(w_B - \epsilon)^{(N-j)}} \right)^{1/j} \tag{A.48}$$

$$\iff k \geq \frac{1}{\epsilon} \left[\left(\frac{w_A^j w_B^{(N-j)}}{(w_B - \epsilon)^{(N-j)}} \right)^{1/j} - w_A \right] \tag{A.49}$$

$$\tag{A.50}$$

□

- Exponential: $u(w_i) = 1 - e^{-\alpha w_i}$

Proof.

$$SW(W') \geq SW(W) \tag{A.51}$$

$$\iff \sum_{i=1}^{j}(1-e^{-\alpha(w_i+k\epsilon)}) + \sum_{i=j+1}^{N}(1-e^{-\alpha(w_i-k)}) \geq \sum_{i=1}^{N}(1-e^{-\alpha(w_i)}) \tag{A.52}$$

$$\iff je^{-\alpha(w_A+k\epsilon)} \leq je^{-\alpha(w_A)} + (N-j)e^{-\alpha(w_B)} - (N-j)e^{-\alpha(w_B-k)} \tag{A.53}$$

$$\iff e^{-\alpha k\epsilon} \leq 1 + \frac{(N-j)}{j}e^{-\alpha(w_B-w_A)}(1-e^{\alpha\epsilon}) \tag{A.54}$$

$$\iff k \geq \left(\frac{-1}{\alpha\epsilon}\right) ln\left[1 + \frac{(N-j)}{j}e^{-\alpha(w_B-w_A)}(1-e^{\alpha\epsilon})\right] \tag{A.55}$$

□

(b) Nash Case:

- Wealth: $u(w_i) = w_i$

  Proof.

$$SW(W') \geq SW(W) \tag{A.56}$$

$$\iff \Pi_{i=1}^{j}(w_i+k\epsilon)\Pi_{i=j+1}^{N}(w_i-\epsilon) \geq \Pi_{i=1}^{N}w_i \tag{A.57}$$

$$\iff (w_A+k\epsilon)^j \geq \frac{w_A^j w_B^{N-j}}{(w_B-\epsilon)^{N-j}} \tag{A.58}$$

$$\iff k \geq \frac{1}{\epsilon}\left[\left(\frac{w_A^j w_B^{N-j}}{(w_B-\epsilon)^{N-j}}\right)^{(1/j)} - w_A\right] \tag{A.59}$$

□

- Polynomial: $u(w_i) = w_i^\alpha$

  Proof.

$$SW(W') \geq SW(W) \tag{A.60}$$
$$\iff \Pi_{i=1}^{j}(w_i + k\epsilon)^\alpha \Pi_{i=j+1}^{N}(w_i - \epsilon)^\alpha \geq \Pi_{i=1}^{N} w_i^\alpha \tag{A.61}$$
$$\iff \Pi_{i=1}^{j}(w_i + k\epsilon)\Pi_{i=j+1}^{N}(w_i - \epsilon) \geq \Pi_{i=1}^{N} w_i \tag{A.62}$$

  which is exactly the same case as for the wealth utility function, and therefore

$$\tag{A.63}$$
$$\iff k \geq \frac{1}{\epsilon}\left[\left(\frac{w_A^j w_B^{N-j}}{(w_B - \epsilon)^{N-j}}\right)^{(1/j)} - w_A\right] \tag{A.64}$$

  □

- Logarithmic: $u(w_i) = ln(w_i)$

Proof.

$$SW(W') \geq SW(W) \tag{A.65}$$

$$\iff \Pi_{i=1}^{j} ln(w_i + k\epsilon)\Pi_{i=j+1}^{N} ln(w_i - \epsilon) \geq \Pi_{i=1}^{N} ln(w_i) \tag{A.66}$$

$$\iff (ln(w_A + k\epsilon))^j \geq \frac{(lnw_A)^j (lnw_B)^{N-j}}{(ln(w_B - \epsilon))^{N-j}} \tag{A.67}$$

$$\iff k \geq \frac{1}{\epsilon} \left\{ exp \left[\frac{(lnw_A)^j (lnw_B)^{N-j}}{(ln(w_B - \epsilon))^{N-j}} \right]^{1/j} - w_a \right\} \tag{A.68}$$

□

- Exponential: $u(w_i) = 1 - e^{-\alpha w_i}$

Proof.

$$SW(W') \geq SW(W) \tag{A.69}$$

$$\iff \Pi_{i=1}^{j}(1-e^{-\alpha(w_i+k\epsilon)})\Pi_{i=j+1}^{N}(1-e^{-\alpha(w_i-\epsilon)}) \geq \Pi_{i=1}^{N}(1-e^{-\alpha w_i}) \tag{A.70}$$

$$\iff (1-e^{-\alpha(w_A+k\epsilon)})^j \geq \frac{(1-e^{-\alpha w_A})^j(1-e^{-\alpha w_B})^{N-j}}{(1-e^{-\alpha(w_B-\epsilon)})^{(N-j)}} \tag{A.71}$$

$$\iff e^{-\alpha(w_A+k\epsilon)} \leq 1-\left[\frac{(1-e^{-\alpha w_A})^j(1-e^{-\alpha w_B})^{N-j}}{(1-e^{-\alpha(w_B-\epsilon)})^{(N-j)}}\right]^{1/j} \tag{A.72}$$

$$\iff k \geq \frac{1}{\epsilon}\left\{\left(\frac{-1}{\alpha}\right) ln\left[1-\left[\frac{(1-e^{-\alpha w_A})^j(1-e^{-\alpha w_B})^{N-j}}{(1-e^{-\alpha(w_B-\epsilon)})^{(N-j)}}\right]^{1/j}\right]-w_A\right\} \tag{A.73}$$

(A.74)

□

Proofs Proposition 3

(a) Robin-hood transfers

We proceed step-by-step by showing that the size of k (Table 2) is well-ordered for each setup in Proposition 3. For example, In order to show that any Robin-hood transfer implying an improvement under Bentham-Wealth is also desired under Bentham-Polynomial, we show that the minimum k required by the former is greater than the minimum k required by the latter.

- B-W implies B-P

  Proof. Considering the values of k indicated on Table 2, we need to show that:

  $$(1/\epsilon)\{[w_1^\alpha + w_2^\alpha - (w_2 - \epsilon)^\alpha]^{1/\alpha} - w_1\} \leq 1 \quad \text{(A.75)}$$

  By rearranging terms and noting that $\epsilon = \lambda(w_2 - w_1)$, $\lambda \in (0, 1)$, since we assume that transfers do not increase inequality, we have:

  $$w_1^\alpha + w_2^\alpha \leq [(1-\lambda)w_1 + \lambda w_2]^\alpha + [(1-\lambda)w_2 + \lambda w_1]^\alpha \quad \text{(A.76)}$$

  Now we note that since w^α is a concave function, it must be the case that:

  $$\lambda_1 w_1^\alpha + (1-\lambda_1)w_2^\alpha \leq [\lambda_1 w_1 + (1-\lambda_1)w_2]^\alpha \quad \text{(A.77)}$$
  $$\lambda_2 w_1^\alpha + (1-\lambda_2)w_2^\alpha \leq [\lambda_2 w_1 + (1-\lambda_2)w_2]^\alpha \quad \text{(A.78)}$$

  where λ_1 , $\lambda_2 \in (0, 1)$

  By setting $\lambda_1 = 1 - \lambda$ and $\lambda_2 = \lambda$, we can sum up the two inequalities above and find:

  $$w_1^\alpha + w_2^\alpha \leq [(1-\lambda)w_1 + \lambda w_2]^\alpha + [(1-\lambda)w_2 + \lambda w_1]^\alpha \quad \text{(A.79)}$$

  This proves our proposition. □

- B-P implies B-L , N-W and N-P

Proof. From Table 2, we must show that:

$$\frac{w_1}{w_2-\epsilon} \leq (1/\epsilon)\{[w_1^{\alpha}+w_2^{\alpha}-(w_2-\epsilon)^{\alpha}]^{1/\alpha}-w_1\} \tag{A.80}$$

By rearranging terms and, once again, taking $\epsilon = \lambda(w_2 - w_1)$, we find:

$$\{(w_1)^{\alpha}+(w_2)^{\alpha}-[(1-\lambda)w_2+\lambda w_1]^{\alpha}\}[(1-\lambda)w_2+\lambda w_1]^{\alpha} \geq w_1^{\alpha}w_2^{\alpha} \tag{A.81}$$

Now define the left term from the inequality as $f(\lambda)$ and note that $f(0) = w_1^{\alpha}w_2^{\alpha}$ and $f(1) = w_1^{\alpha}w_2^{\alpha}$. If we can show that this is a concave function, by continuity it directly follows that the inquality holds for any $\lambda \in (0,1)$ and we have proved the result. By calculating the second derivative, we find:

$$f^{''}(\lambda) = (\alpha^2-\alpha)(w_1^{\alpha}+w_2^{\alpha})(w_1-w_2)^2[(1-\lambda)w_2+\lambda w_1]^{\alpha-2} \leq 0 \tag{A.82}$$

The second derivative is always smaller than zero, since $\lambda \in (0,1)$, $w_2 > w_1 > 0$ and $\alpha \in (0,1)$. This proves our proposition. □

- B-L, N-W and N-P imply N-P

Proof. From Table 2, we must show that:

$$(1/\epsilon)\{e^{[\frac{(lnw_A)^j(lnw_B)^{N-j}}{(lnw_B-\epsilon)^{N-j}}]^{1/j}}-w_A\} \leq \frac{w_1}{w_2-\epsilon} \tag{A.83}$$

By rearranging terms and, once again, taking $\epsilon = \lambda(w_2 - w_1)$, we find:

$$ln\left(\frac{w_1 w_2}{(1-\lambda)w_2 + \lambda w_1}\right) ln[(1-\lambda)w_2 + \lambda w_1] \geq lnw_1 lnw_2 \tag{A.84}$$

Once again, define the left term from the inequality as $f(\lambda)$ and note that $f(0) = lnw_1 lnw_2$ and $f(1) = lnw_1 lnw_2$. If we can show that this is a concave function, by continuity it follows that the inquality holds for any $\lambda \in (0,1)$ and we are done. The second derivative is as follows:

$$f''(\lambda) = -(w_1 - w_2)^2[(1-\lambda)w_2 + \lambda w_1]^{-2} ln\left(\frac{w_1 w_2}{(1-\lambda)w_2 + \lambda w_1}\right) \leq 0 \tag{A.85}$$

Indeed, the second derivative is always smaller than zero, since $\lambda \in (0,1)$, $w_2 > w_1 > 1$ in the Nash-Logarithm form (otherwise, we risk aggregating negative utilities by multiplying them among each other). This proves our proposition. □

- N-P implies Rawls

Proof. In the two groups case, any RH transfer will be desirable under Rawls since it benefits the poorest group. In a more general setup, any RH transfer is weakly preferred, given that it strictly improves welfare when it enriches the least well-off group. Otherwise, in the case in which the individual granted with the transfer is not the least well-off, the new allocation simply bears the same level of welfare under the Rawlsian principle. □

- B-W implies B-E

  Proof. From Table 2, we must show that:

$$(1/\epsilon)\{-(1/\alpha)ln[e^{-\alpha w_1} + e^{-\alpha w_2} - e^{-\alpha(w_2-\epsilon)}] - w_1\} \leq 1 \tag{A.86}$$

  By rearranging terms and, once again, taking $\epsilon = \lambda(w_2 - w_1)$, we find:

$$e^{\alpha[(1-\lambda)w_1+\lambda w_2]}[e^{-\alpha w_1} + e^{-\alpha w_2} - e^{-\alpha[\lambda w_1+(1-\lambda)w_2]}] \geq 1 \tag{A.87}$$

  As before, we define the left term from the inequality as $f(\lambda)$ and note that $f(0) = 1$ and $f(1) = 1$, satisfying the inequality. Subsequently, we study the sign of the derivative to prove our proposition:

$$f'(\lambda) = \alpha(w_2 - w_1)e^{\alpha[(1-\lambda)w_1+\lambda w_2]}[e^{-\alpha w_1} + e^{-\alpha w_2} - 2e^{-\alpha[\lambda w_1+(1-\lambda)w_2]}] \tag{A.88}$$

  We note that $f'(0) = \alpha(w_2 - w_1)e^{\alpha w_1}[e^{-\alpha w_1} - e^{-\alpha w_2}] > 0$, $f'(1) = \alpha(w_2 - w_1)e^{\alpha w_2}[e^{-\alpha w_2} - e^{-\alpha w_1}] < 0$ and that $f'(\lambda)$ sign depends exclusively on the term inside the brackets. This term is strictly decreasing in λ and, therefore, the derivate sign only changes from positive to negative at one point. This implies that the inequality holds since $f(\lambda)$ is continuous and increasing at $\lambda = 0$, has only one maximum point for $\lambda \in (0, 1)$ and satisfies the inequality at $\lambda = 0$ and $\lambda = 1$. This proves our proposition. □

- B-E implies N-E

Proof. We need to show that:

$$\frac{1}{\epsilon}\left\{\left(\frac{-1}{\alpha}\right) ln\left[1-\left[\frac{(1-e^{-\alpha w_1})(1-e^{-\alpha w_2})}{1-e^{-\alpha(w_2-\epsilon)}}\right]\right]-w_1\right\} \leq$$
$$\leq \frac{1}{\epsilon}\left\{\left(\frac{-1}{\alpha}\right) ln[e^{-\alpha w_1}+e^{-\alpha w_2}-e^{-\alpha(w_2-\epsilon)}]-w_1\right\} \tag{A.89}$$

By rearranging terms and, once again, taking $\epsilon = \lambda(w_2 - w_1)$, we find:

$$[e^{-\alpha w_1}+e^{-\alpha w_2}]e^{-\alpha[\lambda w_1+(1-\lambda)w_2]}-e^{-2\alpha[\lambda w_1+(1-\lambda)w_2]}]-e^{-\alpha w_1}e^{-\alpha w_2} \geq 0 \tag{A.90}$$

As before, we define the left term from the inequality as $f(\lambda)$ and note that $f(0) = 0$ and $f(1) = 0$, satisfying the inequality. Subsequently, we study the sign of the derivative to prove our proposition:

$$f^{'}(\lambda) = \alpha(w_2 - w_1)e^{\alpha[\lambda w_1+(1-\lambda)w_2]}[e^{-\alpha w_1}+e^{-\alpha w_2}-2e^{-2\alpha[\lambda w_1+(1-\lambda)w_2]}] \tag{A.91}$$

Note that $f'(0) = \alpha(w_2-w_1)e^{\alpha w_2}[e^{-\alpha w_1}+e^{-\alpha w_2}-2e^{-2\alpha w_2}] > 0$ and that $f'(\lambda)$ sign depends exclusively on the term inside the brackets. This term is strictly decreasing in λ and, therefore, the derivate only changes sign at one point, if ever. This implies that the inequality holds since $f(\lambda)$ is continuous and increasing at $\lambda = 0$, has at most one maximum point for $\lambda \in (0, 1)$ and satisfies the inequality at $\lambda = 0$ and $\lambda = 1$. This proves our proposition. □

- N-E implies Rawls
  The exact same reasoning used to show that N-P implies Rawls also holds here.

(b) Efficiency-Improving Transfers
The proof of proposition 3 for E-I transfers follows analogously to those of the R-H case, with the difference that $w_2 - w_1 < 0$ and ϵ is now bounded by w_2 (an individual cannot be left with negative wealth) instead of $w_2 - w_1$.

First Period Punishment

Treatment	StratP0	StratP5	StratP10	StratP15	StratP20	StratP25	StratP30
Control							
(mean)	1.18	2.07	2.40	2.96	4.13	4.82	5.79
(median)	0	0	2	3	4	5	5
(SD)	3.31	4.09	2.74	2.97	3.94	4.68	5.72
Normative							
	.36	1.73	2.58	3.58	4.29	4.87	5.73
	0	1	2	2	4	5	5
	1.05	2.86	3.13	4.36	4.44	5.03	6.01
Informational							
	1.38	1.90	3.36	3.74	5.14	6.17	7.33
	0	1	3	4	5	6.5	8.5
	3.60	2.16	3.83	3.63	5.00	5.48	6.15
Total	.96	1.90	2.77	3.42	4.51	5.27	6.26
	0	1	2	3	5	5	6
	2.88	3.14	3.25	3.69	4.46	5.06	5.96

Table B.12: Average First Period Punishment by Levels of dictator Taking. StratP0 = dictator take 0 tokens from receiver

Second Period Punishment

Treatment	Punish0	Punish5	Punish10	Punish15	Punish20	Punish25	Punish30
Control							
(mean)	1.02	2.02	2.73	3.36	4.42	5.02	6.22
(median)	0	1	2	2	4	4	6
(SD)	3.18	3.45	3.49	3.58	4.38	5.24	6.26
Normative							
	.62	1.69	2.47	3.18	3.87	4.2	5.18
	0	1	2	3	3	4	4
	1.99	2.86	3.27	3.63	4.31	4.83	6.02
Informational							
	1.5	2.26	2.95	3.60	4.60	5.48	6.31
	0	1	3	4	5	6	8
	4.07	3.92	3.66	3.87	4.24	4.83	5.54
Total							
	1.04	1.98	2.71	3.37	4.29	4.89	5.89
	0	1	2	3	4.5	5	6
	3.16	3.41	3.45	3.67	4.29	4.96	5.93

Table B.13: Average Second Period Punishment by Levels of dictator Taking.Punish0 = dictator take 0 tokens from receiver

Appendix C.

Description of the Variables Used in the Regressions

Table C.14: Variables

Variable	Description
degree	1 if subject completed 8th grade ("scuola media"), 2 if subject completed high school, 3 if subject has a bachelor degree or equivalent, 4 if subject has a master degree or equivalent, 5 if subject has a PhD or equivalent
worker	binomial variable, 1 if worker
male	binomial variable, 1 if male
age	subject's age
social	binomial variable, 1 if subject is a student in social sciences and medicine
arts	binomial variable, 1 if subject is a student in arts or humanities
field_other	binomial variable, 1 if subject not in social or arts
DictatorTake	total amount of tokes a subject when choosing as a dictator takes to the reeiver in the 2 periods
risk	$\in$ [1, 10], 1 if to question "In general, do you consider yourself ready to take risks?" the answer is "Not at all", 10 if the answer is "Totally ready to take risks"
logic	$\in$ [0, 2], 1 point for each correct answer. See figures C.6 and C.7 below for the 2 questions.
impulsivity	$\in$ [0, 3], 1 point for each correct answer. See figures C.8, C.9 and C.10 below for the 3 questions.
NORMATIVE	binomial variable, 1 for subjects in normative treatment

Variable	Description
TREATED	binomial variable, 1 for subject either in normative or in informational treatments
Strat_Punish	punishment first period
Beliefs_Punish	beliefs about peers' average punishment first period
Abs_p0Belifs	absolute value (Strat_Punish - Beliefs_Punish)
Abs Signalp0	absolute vale (Strat_Punish - Peers' average punishment period 1)
Abs BelAvgPun	absolute value (Beliefs_Punish - Peers' average punishment period 1)
NorStratPunish	NORMATIVE*Strat_Punish
NorAbs p0Belifs	NORMATIVE*Abs_p0Belifs
NormAbs_BeliefAvgPunish	NORMATIVE*Abs_BeliefAvgPunish

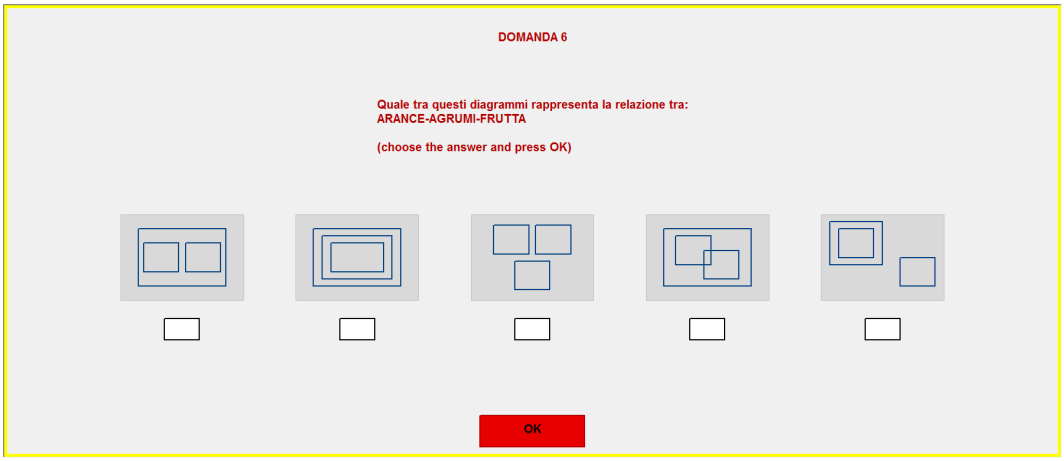


Figure C.6:

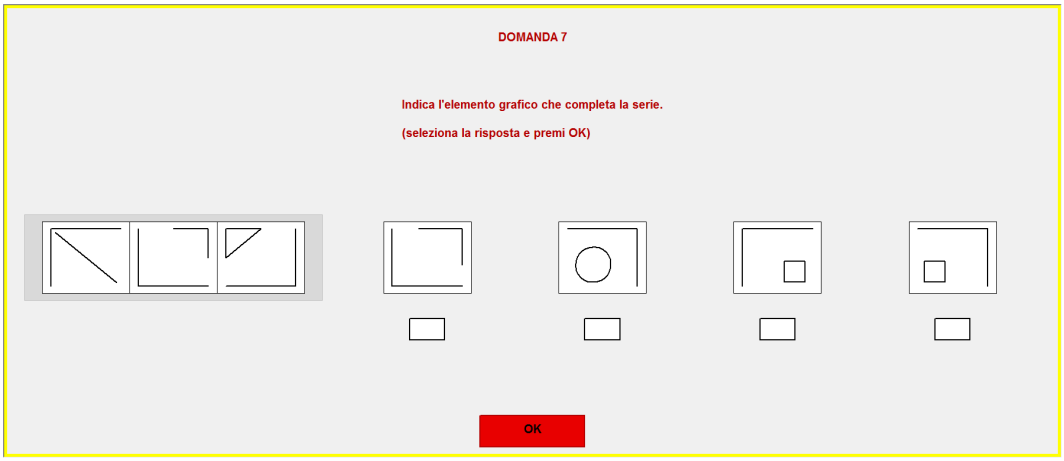


Figure C.7:

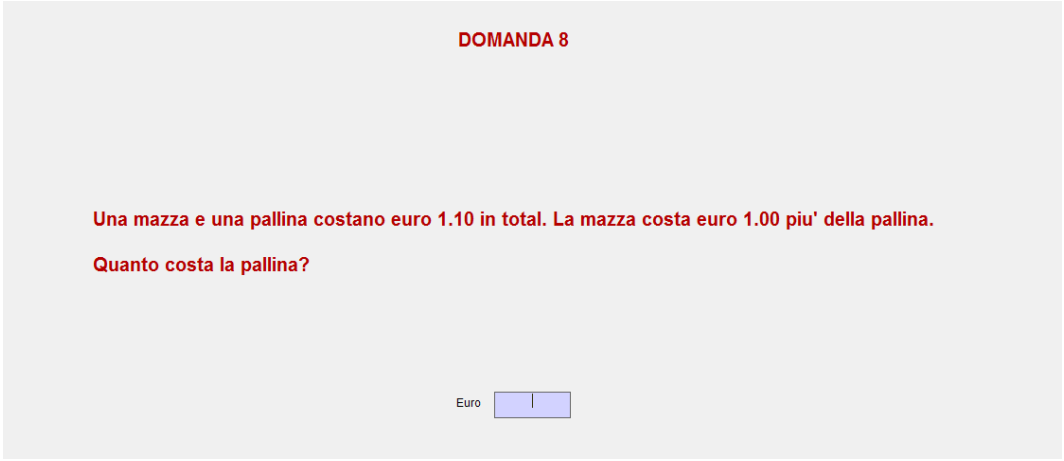


Figure C.8:

Figure C.9:

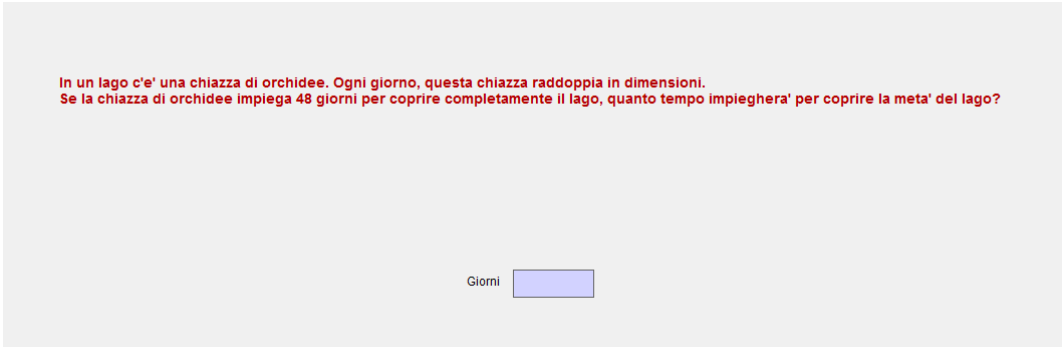


Figure C.10:

English Translation of the Original Italian Instuctions

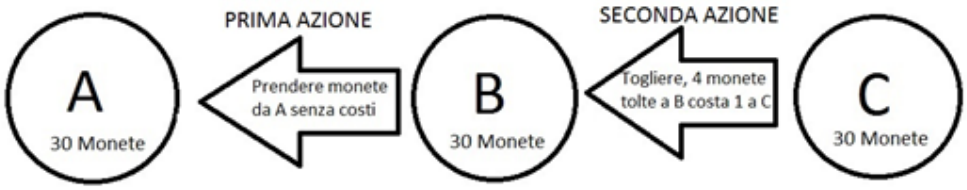
Welcome! This is a study on individual decision-making. Participants' answers are completely anonymous. It will not be possible for data analysts to link individual answers to the participants that provided them. You earned five euro for showing up on time today. Additionally, you can collect other earnings. The amount of these earnings depends on your choices and from the choices other participants will make during the study. During the study you will earn "tokens". For each 10 tokens earned, one euro will be paid out to you. In the unlikely case you will collect negative earnings, those losses will be subtracted from your participation fee. If you have questions at any time, please raise your hand and wait for a researcher that will answer your questions privately. Please switch off and remove from the table any electronic device, do not talk or communicate with other participants during the study. The study is composed of more parts. Earnings obtained in each part of the study are independent from those obtained in the other parts. Your final earnings are composed by:

- Euro 5 of the participation fee
- Earnings collected in the first part of the study
- Earnings collected in one part after the first one. At the end of the study the computer will randomly select the part after the first one of which your earnings will be paid out to you

Final earnings will be paid privately and cash at the end of the study

First Part Instructions: description of the situation (Instructions on this part are the same in the 3 treatments)
Consider a situation with 3 people. Each person is randomly assigned to a role: one "Person A", one "Person B" and one "Person C". A, B and C could make decisions and earn tokens.

- Person A receives 30 tokens and does not make decisions
- Person B receives 30 tokens. Moreover, B could take some or all A's tokens and add them to his own earnings without incurring costs. Precisely, B could take 0, 5, 10, 15, 20, 25 or 30 tokens from A.
- Person C receives 30 tokens, observes B's action and could eliminate some of B's tokens, incurring a cost. For each 4 tokens eliminated from B's earnings, A has to pay 1 token. Person C could use up to 20 tokens to reduce B's earnings. C's decision does not affect A's earnings



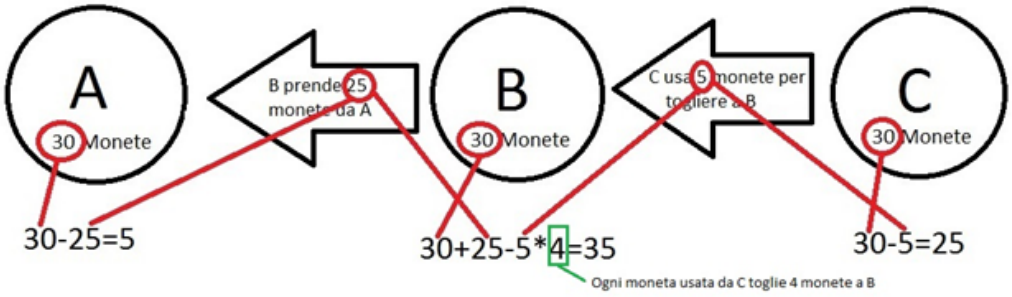
Therefore, A, B and C earnings are:

- Person A: (30 initial tokens) – (tokens taken by B)
- Person B: (30 initial tokens) + (tokens taken from A) – (4*tokens used by C)
- Person C: (30 initial tokens) – (tokens used for reducing B's earnings)

Example 1) (please look at your computer screen): B takes 25 tokens from A. After observing B's choice, C decides to use 5 tokens to reduce B's earnings. Therefore participants' final earnings are:

- Person A = 5 tokens (tokens left by B)

- Person B = 35 tokens (30 initial tokens + 25 tokens taken from A – 5*4= 20 tokens coming from the 5 tokens used by C to reduce B's earnings)
- Person C = 25 tokens (30 initial tokens – 5 tokens used to reduce B earnings)

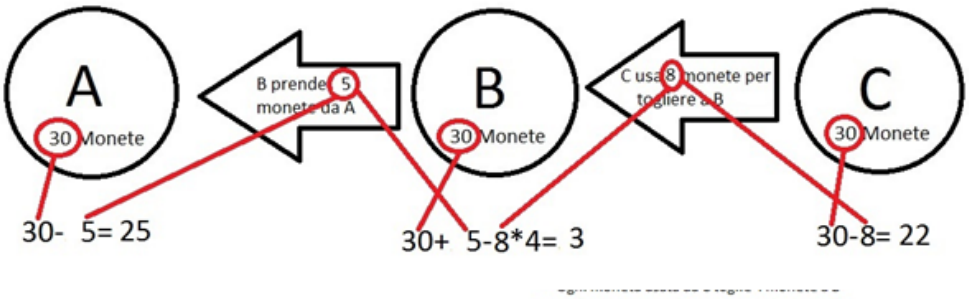


Example 2) (please look at the computer screen): B takes 5 tokens from A. After observing B's choice, C uses 8 tokens to reduce B's earnings. Therefore participants' final earnings are:

- Person A = 25 tokens (left by B)
- Person B = 3 tokens (30 initial tokens + 5 tokens taken from A – 8*4=32 tokens coming from the 8 tokens used by C to reduce B's earnings
- Person C = 22 tokens (30 initial tokens – 8 tokens used to reduce B's earnings)

Your actions and earnings

Person C observes how many tokens B takes from A. You and the other participants in the laboratory have to indicate the number of tokens, an integer between 0 and 20, that C in your opinion will use. When everyone has answered, I calculate the average of the individual

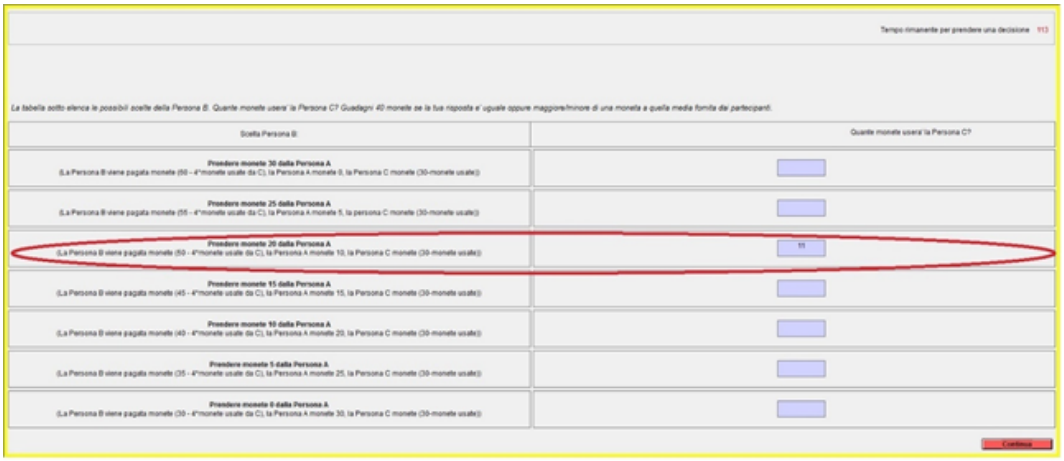


amounts indicated by you and the other participants. If the number you indicated is equal to, or bigger or smaller by one unit than the average, you receive 40 tokens that will be added to your final earnings (if you indicate 0, you will receive the forty tokens if the average is 0, 1 or 2; if you indicate 20, you will receive the 40 tokens if the average is 20, 19 or 18). Instead, you do not earn tokens in this part of the study if the number you indicate is bigger or smaller by more than one unit with respect to the average.

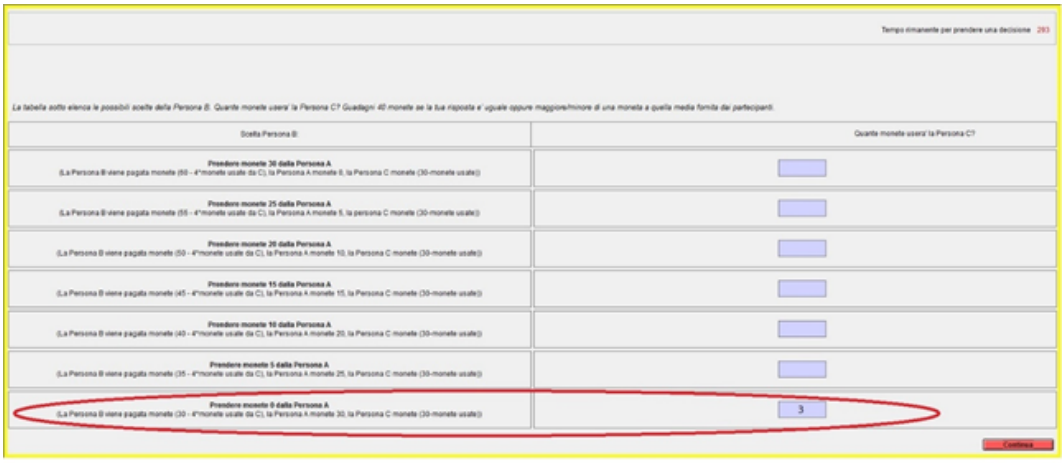
Example 1) (please look at the computer screen): Consider the action of B "take 20 tokens from A and collect 50 tokens, leaving 10 tokens to A". You indicate that C uses 11 tokens. You receive 40 tokens if on average all the participants to the study indicated "11", "10" or "12" tokens. If the average is different from these numbers, you will not earn tokens for this part of the study

Example 2) (please look at the computer screen): Consider the action of B "take 0 tokens from A and collect 30 tokens, leaving 30 tokens to A". You indicate that C uses 3 tokens. You receive 40 tokens if on average all the participants to the study indicated "3", "2" or "4" tokens. If the average is different from these numbers, you will not earn tokens for this part of the study.

You are required to indicate how many tokens Person C uses for each possible action of B (B takes 30 tokens from A; B takes 25 tokens...;

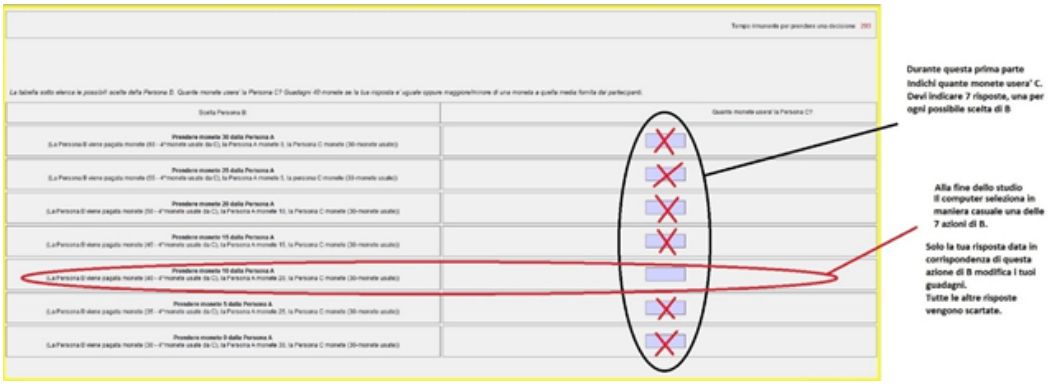


Guadagni le 40 monete se in media i partecipanti allo studio avranno indicato "11" oppure "10" oppure "12". Guadagni 0 se la media e' diversa da questi valori.



Guadagni le 40 monete se in media i partecipanti allo studio avranno indicato "3", "2" oppure "4". Guadagni 0 se la media e' diversa da questi valori

B takes 0 tokens from A). At the end of the study, the computer will randomly select one of the 7 actions of Person B. Relatively to this action, I will verify if you earned the 40 tokens. Your decisions and those of the other participants relative to other possible actions of Person B will be discarded and will not affect your final earnings.



Before starting this first part of the study, I ask you to answer some control questions. Answers to these control questions will not affect your final earnings.
(Participants answer control questions on their computers. The Ztree file containing the control questions is available upon request to the authors).

Instruction second part: description of the situation (instructions on this part are the same in all treatments)

Consider the same situation described in the first part, where 3 people are present, A, B and C, that can make decisions and earn tokens. Exactly as in the first part:

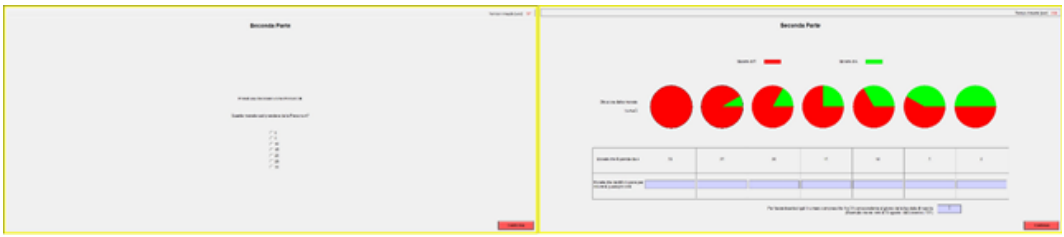
- A receives 30 tokens and does not make decisions
- B receives 30 tokens and could take some or all of the tokens of A

- C receives 30 tokens, observes the action of B and could reduce earnings of B paying a cost (for every 4 tokens of reduction of B's earnings C has to pay 1 token)

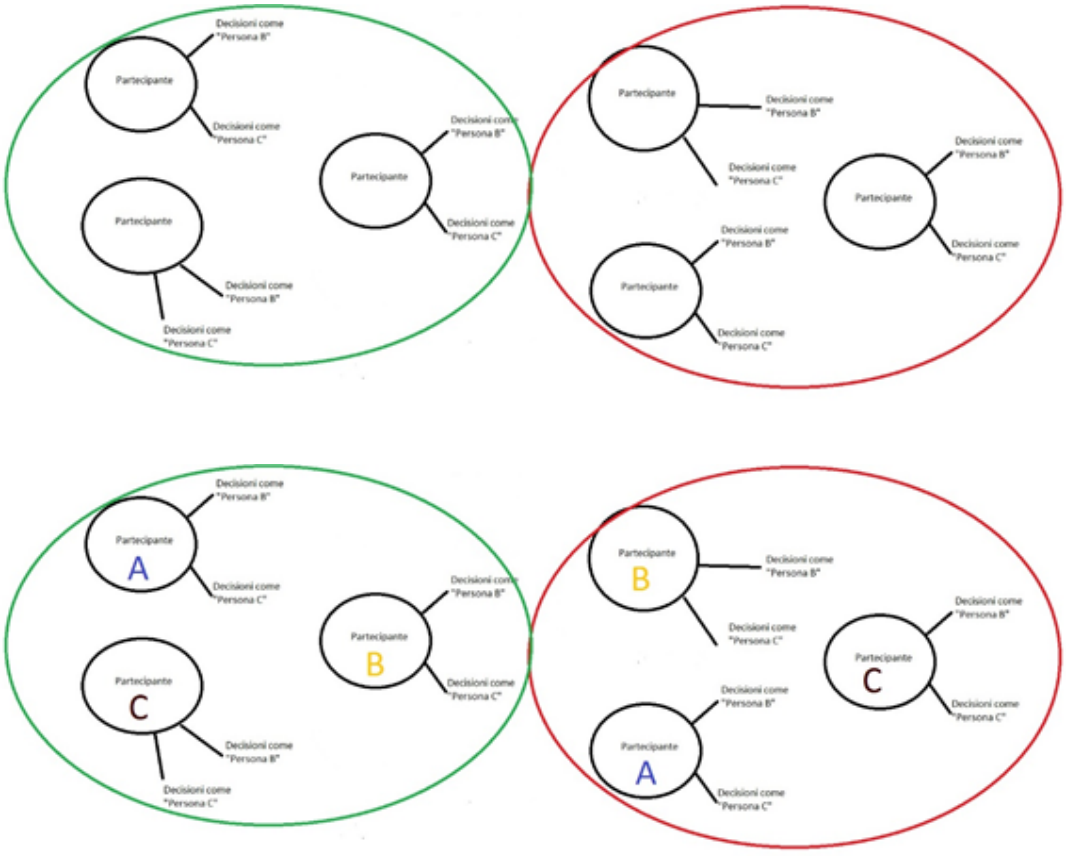
Your actions and earnings

In this second part you and the other participants have to make decisions first as "Person B" then as a "Person C". Therefore, you have to indicate:

- First, as "Person B", how many tokens you take from A
- After, as a "Person C", for any possible action of B how many tokens you use for reducing B's earnings

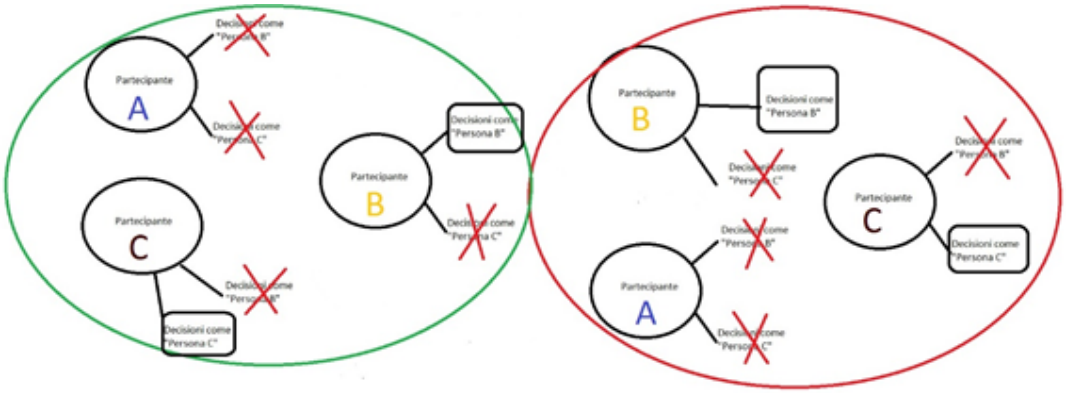


Why do you have to make decisions both as "Person C" and as a "Person B"? In calculating final earnings, each participant is associated to an unique role: either Person A or Person B or Person C. However, you and the other participants will not know which role has been assigned to you until the end of the study today. Indeed, you and the other participants will be randomly divided in groups of 3.
Within the group, each one of the 3 participants is assigned either to role A, B or C.
Assignment to groups and assignment of roles is completely random and each participant has 1 possibility over 3 of being assigned a specific role. Therefore, if you are assigned the role "Person A", your final



earnings are determined by the tokens left you by the Person B that is in your same group. Other decisions you make as a Person B or C will be discarded and have no influence on your final earnings nor on the earnings of the other participants. Similarly, participants assigned to the role "Person B" determine their final earnings and those of the other group components only by the decisions make as Person B. Decisions made as Person C have no effects on final earnings. Finally, also Participants assigned to role "Person C" only influence final earnings only by decisions make as C.

During this second part of the study I will also ask to indicate the day of the month in which you where born (E.g. if you were born January



25th 1983 you should report "25").
Earnings of A, B and C in this second part are determined exactly as in the first part:

- Person A: (30 initial tokens) – (tokens taken by B)
- Person B: (30 initial tokens) + (tokens taken from A) – (4*tokens used by C)
- Person C: (30 initial tokens) – (tokens used for reducing B's earnings)

Before starting this first part of the study, I ask you to answer some control questions. Answers to these control questions will not affect your final earnings.

Instruction third part (Normative Treatment; instructions for Control and Informational are available upon request)

Now the third and last part of this study starts. After the end of this part, I will ask you to fill in a brief questionnaire and then I will proceed with payments. Consider exactly the same situation of the

second part of the study, same roles of A, B and C, same possible decisions that B and C have to make and same initial endowments and possible earnings. As in the second part, you have to make decisions first as a Person B then as a Person C. Additionally, in this third part before making your decisions you will receive information regarding the other participants. You will receive information on decisions made as Person C by the participants at today study. You will know how many tokens on average participants used in the second part of the study to reduce B's earnings. You will receive this information for any of the 7 possible B's choices.

Furthermore, before the end of the study, individual decisions as "Person C" that you are going to make in this third part will be revealed to 5 participants randomly selected. Similarly, you will received information regarding the individual choices made as Person C by 5 other participants

Each participant will be randomly assigned to an ID number. The ID number assigned is independent from the number of the PC you sit on. After you saw the individual choices of the other 5 participants, you and the other participants will be able to vote for sending a smiling or

Monete che B prende da A	**Media** monete riduzione guadagni B **parte 2**	Monete usate **Partecipante 1** parte 3	Monete usate **Partecipante 2** parte 3	Monete usate **Partecipante 3** parte 3	Monete usate **Partecipante 4** parte 3	Monete usate **Partecipante 5** parte 3
0	0	0	0	0	0	0
5	0	0	0	0	0	0
10	0	0	0	0	0	0
15	0	0	0	0	0	0
20	0	0	0	0	0	0
25	0	0	0	0	0	0
30	0	0	0	0	0	0

sad emoticon

Partecipante 15	☐ Sorridente	☐ Triste
Partecipante 16	☐ Sorridente	☐ Triste
Partecipante 17	☐ Sorridente	☐ Triste

You receive a smiling emoticon if the majority of the five participants that saw your choices vote for "smiling". Otherwise you will receive a sad emoticon. The emoticon will remain on your screen for one minute, then disappears automatically. After this minute has passed, you will know your final earnings.

If you have questions, please raise your hand and I will answer to you privately. Otherwise push the "Continue" button and start with the

third part.

References

Akerlof, G. A., 1980. A theory of social custom, of which unemployment may be one consequence. The quarterly journal of economics 94 (4), 749–775.

Akerlof, G. A., 1989. The economic of illusion. Economics & Politics 1 (1), 1–15.

Akerlof, G. A., Shiller, R. J., 2010. Animal spirits: How human psychology drives the economy, and why it matters for global capitalism. Princeton University Press.

Akerlof, G. A., Yellen, J. L., Katz, M. L., 1996. An analysis of out-of-wedlock childbearing in the united states. The Quarterly Journal of Economics 111 (2), 277–317.

Allingham, M. G., Sandmo, A., 1972. Income Tax Evasion: a Theoretical Analysis. Journal of Public Economics 1 (3-4), 323–338.

Alm, J., McClelland, G. H., Schulze, W. D., 1992. Why Do People Pay Taxes? Journal of Public Economics 48 (1), 21–38.

Ameriks, J., Caplin, A., Leahy, J., Tyler, T., 2007. Measuring self-control problems. The American Economic Review, 966–972.

Amir, O., Ariely, D., Cooke, A., Dunning, D., Epley, N., Gneezy, U., Koszegi, B., Lichtenstein, D., Mazar, N., Mullainathan, S., et al., 2005. Psychology, behavioral economics, and public policy. Marketing Letters 16 (3-4), 443–454.

Andersen, S., Harrison, G. W., Lau, M. I., Rutström, E. E., 2008. Eliciting risk and time preferences. Econometrica 76 (3), 583–618.

Andreoni, J., Erard, B., Feinstein, J., 1998. Tax Compliance. Journal of economic literature 36 (2), 818–860.

Ariely, D., Jones, S., 2008. Predictably irrational. HarperCollins New York.

Arrow, K. J., 1951. Social Choice and Individual Values. New York: John Wiley and Sons.

Arrow, K. J., 1973. Some Ordinalist-utilitarian Notes on Rawls's Theory of Justice. Journal of Philosophy 70 (9), 245–263.

Asch, S., 1951. Effects of Group Pressure upon the Modification and Distortion of Judgments.

Asch, S., 1956. Studies of Independence and Conformity: A minority of One Against a Unanimous Majority. Psychological Monographs: General and Applied 70 (9), 1–70.

Aviram, A., 2004. A Paradox of Spontaneous Formation: The Evolution of Private Legal Systems. Yale Law and Policy Review 22, 1–65.

Axelrod, R., 1986. An Evolutionary Approach to Norms. American political science review 80 (04), 1095–1111.

Baron, J., 1993. Morality and rational choice. Vol. 18. Dordrecht (NL): Kluwer Academic Publisher.

Battiston, P., Gamba, S., 2013. Is Tax Compliance a Social Norm? A Field Experiment. Working paper, University of Milano-Bicocca, Department of Economics.

Becker, G., 1976. Toward a more general theory of regulation. Journal of Law and Economics, 245–248.

Becker, G. S., 1968a. Crime and punishment: An economic approach. Journal of Political Economy 76 (2), 169–217.

Becker, G. S., 1968b. Crime and Punishment: An Economic Approach. The Journal of Political Economy 76 (2), 169–217.

Becker, G. S., 1991. A note on restaurant pricing and other examples of social influences on price. Journal of Political Economy 99 (5), 1109.

Becker, G. S., Becker, G. S., 2009. A Treatise on the Family. Harvard university press.

Benabou, R., Tirole, J., 2003. Intrinsic and extrinsic motivation. The Review of Economic Studies 70 (3), 489–520.

Bendor, J., Mookherjee, D., 1990. Norms, Third-party Sanctions, and Cooperation. JL Econ & Org. 6, 33.

Benhabib, J., Bisin, A., Schotter, A., 2010. Present-bias, quasi-hyperbolic discounting, and fixed costs. Games and Economic Behavior 69 (2), 205–223.

Bentham, J., 1776. A Fragment on Government: Being an Examination of what is Delivered, on the Subject of Government in General, in the Introduction to Sir William Blackstone's Commentaries: with a Preface, in which is Given a Critique on the Work at Large. Edited by T. Payne, P. Elmsly, and E. Brooke.

Bergson, A., 1938. A reformulation of certain aspects of welfare economics. The Quarterly Journal of Economics 52 (2), 310–334.

Bernhard, H., Fischbacher, U., Fehr, E., 2006. Parochial Altruism in Humans. Nature 442 (7105), 912–915.

Bernheim, B., 1994. A Theory of Conformity. Journal of political Economy, 841–877.

Bernheim, B. D., Rangel, A., 2004. Addiction and cue-triggered decision processes. American Economic Review, 1558–1590.

Bernheim, B. D., Rangel, A., 2005. Behavioral public economics: Welfare and policy analysis with non-standard decision-makers. Tech. rep., National Bureau of Economic Research.

Bernheim, B. D., Rangel, A., 2012. Behavioral Public Economics: Welfare and Policy Analysis with Nonstandard Decision-Makers. Princeton: Princeton University Press.

Berton, P., 1998. How Unique is Japanese Negotiating Behavior? Japan Review 10, 151–161.

Bertrand, M., Mullainathan, S., Shafir, E., 2006. Behavioral economics and marketing in aid of decision making among the poor. Journal of Public Policy & Marketing 25 (1), 8–23.

Bertsimas, D., Farias, V. F., Trichakis, N., 2011. The price of fairness. Operations research 59 (1), 17–31.

Beshears, J., Weller, B., 2010. Public policy and saving for retirement: The "autosave" features of the pension protection act of 2006. Better living through economics, 274–288.

Binmore, K., 1989. Social contract i: Harsanyi and rawls. The Economic Journal 99 (395), 84–102.

Birch, S., Melnikow, J., Kuppermann, M., 2003. Conservative versus aggressive follow up of mildly abnormal pap smears: testing for process utility. Health Economics 12 (10), 879–884.

Bird, R., 1992. Tax reform in latin america. Latin American Research Review 27 (1), 7–36.

Boadway, R. W., Bruce, N., 1984. Welfare economics. Blackwell Oxford.

Bond, R., Smith, P., 1996. Culture and Conformity: A Meta-analysis of Studies using Asch's (1952b, 1956) Line Judgment Task. Psychological Bulletin; Psychological Bulletin 119 (1), 111.

Bonezzi, A., Brendl, C. M., De Angelis, M., 2011. Stuck in the middle the psychophysics of goal pursuit. Psychological science 22 (5), 607–612.

Bose, P., 1995. Regulatory Errors, Optimal Fines and the Level of Compliance. Journal of Public Economics 56 (3), 475–484.

Brouwer, W. B., Culyer, A. J., Van Exel, N., Rutten, F. F., 2008. Welfarism vs. extra-welfarism. Journal of health economics 27 (2), 325–338.

Bruni, L., 2007. Handbook on the Economics of Happiness. Edward Elgar Publishing.

Buchanan, J. M., 1959. Positive economics, welfare economics, and political economy. JL & Econ. 2, 124.

Buchanan, J. M., Brennan, G., Tollison, R. D., 1979. What should economists do? Liberty Press Indianapolis.

Buckholtz, J., Asplund, C., Dux, P., Zald, D., Gore, J., Jones, O., Marois, R., 2008. The Neural Correlates of Third-party Punishment. Neuron 60 (5), 930–940.

Burnkrant, R. E., Cousineau, A., 1975. Informational and normative social influence in buyer behavior. Journal of Consumer research, 206–215.

Burrows, P., 1995. Analyzing legal paternalism. International Review of Law and Economics 15 (4), 489–508.

Calabresi, G., 1970. The cost of accidents: a legal and economic analysis. New Haven: Yale University Press.

Calabresi, G., 1985. Ideals, Beliefs, Attitudes, and the Law: Private Law Perspectives on a Public Law Problem. New York: Syracuse University Press.

Camerer, C., Issacharoff, S., Loewenstein, G., O'donoghue, T., 2003. Regulation for conservatives: Behavioral economics and the case for asymmetric paternalism. University of Pennsilvanya Law Review 151, 1211.

Camerer, C., Loewenstein, G., 2004a. Behavioral economics: Past, present, future. Princeton: Princeton University Press.

Camerer, C. F., Loewenstein, G., 2004b. Behavioral Economics: Past, Present, Future. In: Camerer, C. F., Loewenstein, G., Rabin, M. (Eds.), Advances in behavioral economics. Princeton University Press, Princeton, pp. 3–51.

Camerer, C. F., Loewenstein, G., Rabin, M., 2004. Advances in Behavioral Economics. Princeton University Press, Princeton.

Carbonara, E., Parisi, F., von Wangenheim, G., 2012. Unjust Laws and Illegal Norms. International Review of Law and Economics 32 (3), 285–299.

Carpenter, J., 2007. The Demand for Punishment. Journal of Economic Behavior & Organization 62 (4), 522–542.

Carpenter, J., Holmes, J., Matthews, P. H., 2010. Charity Auctions in the Experimental Lab. Research in Experimental Economics 13, 201–249.

Carroll, G. D., Choi, J. J., Laibson, D., Madrian, B. C., Metrick, A., 2009. Optimal defaults and active decisions. The quarterly journal of economics 124 (4), 1639–1674.

Casal, S., Mittone, L., 2014. Social esteem versus social stigma: the role of anonymity in an income reporting game. CEEL Working Papers 1401, Cognitive and Experimental Economics Laboratory, Department of Economics, University of Trento, Italia.

Casari, M., Luini, L., 2009. Cooperation under Alternative Punishment Institutions: An experiment. Journal of Economic Behavior & Organization 71 (2), 273–282.

Cason, T. N., Mui, V.-L., 1998. Social influence in the sequential dictator game. Journal of Mathematical Psychology 42 (2), 248–265.

Chabris, C. F., Simons, D. J., 2011. The invisible gorilla: And other ways our intuitions deceive us. Random House LLC.

Chang, C.-J., Kuo, H.-C., Chen, C.-Y., Chen, T.-H., Chung, P.-Y., 2012. Ergonomic Techniques for a Mobile E-invoice System: Operational Requirements of an Information Management System. Human Factors and Ergonomics in Manufacturing Service Industries, n/a–n/a.
URL http://dx.doi.org/10.1002/hfm.20340

Chang, H. F., 2000. A liberal theory of social welfare: fairness, utility, and the pareto principle. The Yale Law Journal 110 (2), 173–235.

Chang, J.-j., Lai, C.-c., 2004. Collaborative Tax Evasion and Social Norms: Why Deterrence Does Not Work. Oxford Economic Papers 56 (2), 344–368.

Choi, J. J., Laibson, D., Madrian, B. C., 2011. $100 bills on the sidewalk: Suboptimal investment in 401 (k) plans. Review of Economics and Statistics 93 (3), 748–763.

Choi, J. J., Laibson, D., Madrian, B. C., Metrick, A., 2003. Optimal defaults. American Economic Review, 180–185.

Christakis, N. A., Fowler, J. H., 2007. The spread of obesity in a large social network over 32 years. New England journal of medicine 357 (4), 370–379.

Cialdini, R. B., 1993. Influence: the Psychology of Persuasion. Collins, NY.

Cialdini, R. B., Trost, M. R., 1998. Social influence: Social norms, conformity and compliance. McGraw-Hill.

Coffman, L., 2011. Intermediation Reduces Punishment (and Reward). American Economic Journal: Microeconomics 3 (4), 77–106.

Cohen, J. B., Golden, E., 1972. Informational social influence and product evaluation. Journal of Applied Psychology 56 (1), 54.

Coleman, S., 1996. The minnesota income tax compliance experiment: State tax results.

Conly, S., 2012. Against autonomy: justifying coercive paternalism. Cambridge University Press.

Cooper, D., Rege, M., 2008. Social Interaction Effects and Choice under Uncertainty: an Experimental Study. Tech. rep., University of Stavanger.

Cooter, R., 1998. Expressive law and economics. The Journal of Legal Studies 27 (2), 585–608.

Corazzini, L., Faravelli, M., Stanca, L., 2010. A Prize To Give For: An Experiment on Public Good Funding Mechanisms. The Economic Journal 120 (547), 944–967.

Cowell, F. A., 1990. Tax sheltering and the cost of evasion. Oxford Economic Papers 42 (1), 231–243.

Cox, J. C., Friedman, D., Gjerstad, S., 2007. A tractable model of reciprocity and fairness. Games and Economic Behavior 59 (1), 17–45.

Craswell, R., 2003. Kaplow and shavell on the substance of fairness. The journal of legal studies 32 (1), 245–275.

Culyer, A. J., 1971. Medical care and the economics of giving. Economica, 295–303.

Culyer, A. J., 1989. The normative economics of health care finance and provision. Oxford review of economic policy 5 (1), 34–58.

Davis, D. D., Millner, E. L., Reilly, R. J., 2005. Subsidy schemes and charitable contributions: a closer look. Experimental economics 8 (2), 85–106.

Della Vigna, S., 2009. Psychology and economics: Evidence from the field. Journal of Economic Literature, 315–372.

Deutsch, M., Gerard, H. B., 1955. A study of normative and informational social influences upon individual judgment. The journal of abnormal and social psychology 51 (3), 629.

Devenow, A., Welch, I., 1996. Rational Herding in Financial Economics. European Economic Review 40 (3), 603–615.

Diamond, P. A., 1967. Cardinal welfare, individualistic ethics, and interpersonal comparison of utility: Comment. The Journal of Political Economy 75 (5), 765.

Dinner, I., Johnson, E. J., Goldstein, D. G., Liu, K., 2011. Partitioning default effects: why people choose not to choose. Journal of Experimental Psychology: Applied 17 (4), 332.

Dorff, M., 2002. Why welfare depends on fairness: A reply to kaplow & shavell. Southern California Law Review 75, 847.

Duffy, J., Matros, A., 2012. All-Pay Auctions vs. Lotteries as Provisional Fixed-Prize Fundraising Mechanisms: Theory and Evidence. Working paper, University of Pittsburgh.

Duflo, E., Kremer, M., Robinson, J., 2011. Nudging farmers to use fertilizer: Theory and experimental evidence from kenya. American Economic Review 101, 2350–2390.

Duncan, G., 2013. Should happiness-maximization be the goal of government? In: The Exploration of Happiness. Springer, pp. 303–320.

Dworkin, G., 2010. Paternalism. In: Zalta, E. N. (Ed.), Stanford Encyclopedia of Philosophy. Stanford University.

Ela, J. S., 2008. Law and norms in collective action: Maximizing social influence to minimize carbon emissions. UCLA Journal of Environmental Law & Policy 27 (1).

Ellickson, R., 1999. The Evolution of Social Norms: a Perspective from the Legal Academy. Yale Law School, Program for Studies in Law, Economics and Public Policy, Working Paper (230).

Engel, C., 2011. Dictator games: a meta study. Experimental Economics 14 (4), 583–610.

Falk, A., Fischbacher, U., 2002. "Crime" in the Lab-Detecting Social Interaction. European Economic Review 46 (4), 859–869.

Falk, A., Fischbacher, U., Gächter, S., 2010. Living in Two Neighborhoods—Social Interaction Effects in the Laboratory. Economic Inquiry.

Falk, A., Ichino, A., 2006. Clean Evidence on Peer Effects. Journal of Labor Economics 24 (1), 39–57.

Falkinger, J., Walther, H., 1991. Rewards versus Penalties: on a new Policy against Tax Evasion. Public Finance Review 19 (1), 67–79.

Faravelli, M., Stanca, L., 2012. Single Versus Multiple-prize All-pay Auctions to Finance Public Goods: An Experimental Analysis. Journal of Economic Behavior and Organization 81 (2), 677–688.

Fehr, E., Fischbacher, U., 2004a. Social norms and human cooperation. Trends in cognitive sciences 8 (4), 185–190.

Fehr, E., Fischbacher, U., 2004b. Third-party Punishment and Social Norms. Evolution and human behavior 25 (2), 63–87.

Fehr, E., Fischbacher, U., et al., 2003. The Nature of Human Altruism. Nature 425 (6960), 785–791.

Fehr, E., Gächter, S., 2002. Altruistic Punishment in Humans. Nature 415, 137–140.

Fehr, E., Schmidt, K. M., 1999. A theory of fairness, competition, and cooperation. Quarterly journal of Economics, 817–868.

Feinberg, J., Feinberg, J., 1989. Harm to self. Oxford University Press New York.

Feld, L., Frey, B., Torgler, B., 2006. Rewarding Honest Taxpayers. In: Elffers, H., Verboon, P., Huisman, W. (Eds.), Managing and Maintaining Compliance. BJu Legal Publishers, The Hague, pp. 45–61.

Fischbacher, U., 2007. z-Tree: Zurich Toolbox for Ready-made Economic Experiments. Experimental Economics 10 (2), 171–178.

Fischman, J. B., 2013. Reuniting'is' and'ought'in empirical legal scholarship. University of Pennsylvania Law Review 162 (1), 117.

Fleurbaey, M., Hammond, P. J., 2004. Interpersonally comparable utility. In: Handbook of utility theory. Springer, pp. 1179–1285.

Fleurbaey, M., Tungodden, B., Chang, H. F., 2003. Any non-welfarist method of policy assessment violates the pareto principle: A comment. Journal of Political Economy 111 (6), 1382–1385.

Fortin, B., Lacroix, G., Villeval, M., 2007. Tax Evasion and Social Interactions. Journal of Public Economics 91 (11), 2089–2112.

Foster, J. E., Sen, A., 1997. On economic inequality. Expanded edition. Oxford: Clarendon Press.

Fowler, J. H., 2005. Altruistic punishment and the origin of cooperation. Proceedings of the National Academy of Sciences of the United States of America 102 (19), 7047–7049.

Frey, B. S., Jegen, R., 2001. Motivation Crowding Theory. Journal of Economic Surveys 15 (5), 589–611.

Frey, B. S., Stutzer, A., 2010. Happiness and economics: How the economy and institutions affect human well-being. Princeton University Press.

Friedman, D., 2008. Libertarianism. In: Durlauf, Steven N., B. L. E. (Ed.), The New Palgrave Dictionary of Economics. Palgrave Macmillan.

Friedman, M., Friedman, R., 1990. Free to choose: A personal statement. Houghton Mifflin Harcourt.

Fuster, A., Meier, S., 2010. Another Hidden Cost of Incentives: The Detrimental Effect on Norm Enforcement. Management Science 56 (1), 57–70.

Gächter, S., Fehr, E., 2000. Cooperation and Punishment in Public Goods Experiments. American Economic Review 90 (4), 980–994.

Galbiati, R., Zanella, G., 2012. The tax evasion social multiplier: Evidence from italy. Journal of Public Economics 96 (5), 485–494.

Gauthier, D. P., 1963. Practical reasoning: The structure and foundations of prudential and moral arguments and their exemplification in discourse. Oxford: Clarendon Press.

Giebe, T., Schweinzer, P., 2013. Consuming your Way to Efficiency: Public-goods Provision through non-distorsionary Tax Lotteries. Working paper.

Glaeser, E. L., 2006. Paternalism and psychology. University of Chicago law review 73 (1), 133–156.

Glaeser, E. L., Sacerdote, B., Scheinkman, J. A., 1996. Crime and social interactions. The Quarterly Journal of Economics 111 (2), 507–548.

Gray, J. A., 1981. A Critique of Eysenck's Theory of Personality. In: A Model for Personality. New York: Springer, pp. 246–276.

Greiner, B., 2004. An Online Recruitment System for Economic Experiments.

Gul, F., Pesendorfer, W., 2001. Temptation and self-control. Econometrica 69 (6), 1403–1435.

Gul, F., Pesendorfer, W., 2004. Self-control and the theory of consumption. Econometrica 72 (1), 119–158.

Güth, W., Schmittberger, R., Schwarze, B., 1982. An Experimental Analysis of Ultimatum Bargaining. Journal of Economic Behavior & Organization 3 (4), 367–388.

Hammond, P. J., 1991. Interpersonal comparisons of utility: Why and how they are and should be made. In: Elster, J., Roemer, J. E. (Eds.), Interpersonal comparisons of well-being. Cambridge and New York: Cambridge University Press, pp. 200–254.

Harrison, G. W., Lau, M. I., Rutström, E. E., 2007. Estimating risk attitudes in denmark: A field experiment*. The Scandinavian Journal of Economics 109 (2), 341–368.

Harsanyi, J. C., 1977. Rational behavior and bargaining equilibrium in games and social situations. New York: Cambridge University Press.

Harsanyi, J. C., 1980. Rule utilitarianism, rights, obligations and the theory of rational behavior. Theory and Decision 12 (2), 115–133.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J., Gurven, M., Gwako, E., Henrich, N., et al., 2006. Costly Punishment across Human Societies. Science 312 (5781), 1767–1770.

Herrnstein, R. J., Loewenstein, G. F., Prelec, D., Vaughan, W., 1993. Utility maximization and melioration: Internalities in individual choice. Journal of behavioral decision making 6 (3), 149–185.

Hetcher, S., 2004. Norms in a Wired World. Cambridge University Press.

Higgins, S. T., Wong, C. J., Badger, G. J., Ogden, D. E. H., Dantona, R. L., 2000. Contingent reinforcement increases cocaine abstinence during outpatient treatment and 1 year of follow-up. Journal of Consulting and Clinical Psychology 68 (1), 64.

Hirshleifer, D., Hong Teoh, S., 2003. Herd Behaviour and Cascading in Capital Markets: a Review and Synthesis. European Financial Management 9 (1), 25–66.

Hoff, K., Kshetramade, M., Fehr, E., 2011. Caste and Punishment: the Legacy of Caste Culture in Norm Enforcement. The Economic Journal 121 (556), F449–F475.

Holt, C. A., Laury, S. K., 2002. Risk aversion and incentive effects. The American Economic Review 92 (5), 1644–1655.

Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., Camerer, C. F., 2005. Neural systems responding to degrees of uncertainty in human decision-making. Science 310 (5754), 1680–1683.

Hull, C. L., 1932. The goal-gradient hypothesis and maze learning. Psychological Review 39 (1), 25.

Jeffery, R. W., Gerber, W. M., Rosenthal, B. S., Lindquist, R. A., 1983. Monetary contracts in weight control: effectiveness of group and individual contracts of varying size. Journal of Consulting and Clinical Psychology 51 (2), 242.

Johnson, E. j., Goldstein, D. G., 2013. Decisions by Default. In: Shafir, E. (Ed.), The Behavioral Foundations of Public Policy. Princeton NJ: Princeton University Press, pp. 417–427.

Johnson, E. J., Hershey, J., Meszaros, J., Kunreuther, H., 1993. Framing, probability distortions, and insurance decisions. Springer.

Kahan, D. M., 1997. Social influence, social meaning, and deterrence. Virginia Law Review, 349–395.

Kahneman, D., 2003. Maps of bounded rationality: Psychology for behavioral economics. American economic review, 1449–1475.

Kahneman, D., 2011. Thinking, fast and slow. Macmillan.

Kahneman, D., Krueger, A. B., 2006. Developments in the measurement of subjective well-being. The journal of economic perspectives 20 (1), 3–24.

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., Stone, A. A., 2004. A survey method for characterizing daily life experience: The day reconstruction method. Science 306 (5702), 1776–1780.

Kahneman, D., Tversky, A., 1979. Prospect theory: An analysis of decision under risk. Econometrica: Journal of the Econometric Society, 263–291.

Kahneman, D., Wakker, P. P., Sarin, R., 1997. Back to bentham? explorations of experienced utility. The Quarterly Journal of Economics 112 (2), 375–406.

Kalai, E., Smorodinsky, M., 1975. Other solutions to nash's bargaining problem. Econometrica: Journal of the Econometric Society, 513–518.

Kaplow, L., Shavell, S., 1999. The conflict between notions of fairness and the pareto principle. American Law and Economics Review 1 (1), 63–77.

Kaplow, L., Shavell, S., 2001. Any non-welfarist method of policy assessment violates the pareto principle. Journal of Political Economy 109 (2), 281–286.

Kaplow, L., Shavell, S., 2009. Fairness versus welfare. Cambridge: Harvard university press.

Karlan, D., McConnell, M., Mullainathan, S., Zinman, J., 2010. Getting to the top of mind: How reminders increase saving. Tech. rep., National Bureau of Economic Research.

Karoshi, B., 2008. How to Win the Lottery...by not Playing. Brian S. Mangam, Lulu.com.

Khodadadi, A., Tütüncü, R. H., Zangari, P. J., 2006. Optimisation and quantitative investment management. Journal of Asset Management 7 (2), 83–92.

Kirchler, E., 2007. The Economic Psychology of Tax Behaviour. Cambridge University Press, Cambridge.

Kivetz, R., Urminsky, O., Zheng, Y., 2006. The goal-gradient hypothesis resurrected: Purchase acceleration, illusionary goal progress, and customer retention. Journal of Marketing Research 43 (1), 39–58.

Klick, J., Mitchell, G., 2006. Government regulation of irrationality: Moral and cognitive hazards. Minnesota Law Review 90, 1620.

Klick, J., Parisi, F., 2008. Social networks, self-denial, and median preferences: Conformity as an evolutionary strategy. The Journal of Socio-Economics 37 (4), 1319–1327.

Knight, J., 1998. Justice and fairness. Annual Review of Political Science 1 (1), 425–449.

Köbberling, V., 2006. Strength of preference and cardinal utility. Economic Theory 27 (2), 375–391.

Krupka, E., Weber, R., 2009. The Focusing and Informational Effects of Norms on Pro-social Behavior. Journal of Economic Psychology 30 (3), 307–320.

Krupka, E. L., Weber, R. A., 2013. Identifying social norms using coordination games: Why does dictator game sharing vary? Journal of the European Economic Association 11 (3), 495–524.

Kurzban, R., DeScioli, P., O'Brien, E., 2007. Audience Effects on Moralistic Punishment. Evolution and Human behavior 28 (2), 75–84.

Laibson, D., 1997. Golden eggs and hyperbolic discounting. The Quarterly Journal of Economics 112 (2), 443–478.

Laibson, D., Repetto, A., Tobacman, J., 2007. Estimating discount functions with consumption choices over the lifecycle. Tech. rep., National Bureau of Economic Research.

Landry, C. E., Lange, A., List, J. A., Price, M. K., Rupp, N. G., 2006. Toward an Understanding of the Economics of Charity: Evidence from a Field Experiment. The Quarterly Journal of Economics 121 (2), 747–782.

Lange, A., List, J. A., Price, M. K., 2007. Using Lotteries to Finance Public Goods: Theory and Experimental Evidence. International Economic Review 48 (3), 901–927.

Larsen, R. J., Ketelaar, T., 1991. Personality and Susceptibility to Positive and Negative Emotional States. Journal of Personality and Social Psychology 61 (1), 132.

Layard, P. R. G., Layard, R., 2011. Happiness: Lessons from a new science. Penguin UK.

Le Grand, J., 1997. Knights, knaves or pawns? human behaviour and social policy. Journal of Social Policy 26 (2), 149–169.

Ledyard, J., 1995. Public Goods: A Survey of Experimental Research. In: Kagel, J. H., Roth, A. E. (Eds.), The Handbook of Experimental Economics. Princeton University Press, Princeton.

Lewisch, P., Ottone, S., Ponzano, F., 2011. Free-Riding on Altruistic Punishment? An Experimental Comparison of Third-Party. Review of Law and Economics 7 (1).

Liberman, V., Samuels, S. M., Ross, L., 2004. The name of the game: Predictive power of reputations versus situational labels in determining prisoner's dilemma game moves. Personality and social psychology bulletin 30 (9), 1175–1185.

Lieberman, D., Linke, L., 2007. The Effect of Social Category on Third-party Punishment. Evolutionary Psychology 5 (2), 289–305.

Lin, C., 1992. An appraisal of business tax reform in taiwan: The case of value-added taxation. In: The Political Economy of Tax Reform, NBER-EASE Volume 1. Chicago: University of Chicago Press, pp. 137–155.

List, J. A., 2007. On the interpretation of giving in dictator games. Journal of Political Economy 115 (3), 482–493.

Loewenstein, G., 1996. Out of control: Visceral influences on behavior. Organizational behavior and human decision processes 65 (3), 272–292.

Loewenstein, G., Brennan, T., Volpp, K. G., 2007. Asymmetric paternalism to improve health behaviors. Jama 298 (20), 2415–2417.

Loewenstein, G., Haisley, E., 2007. The economist as therapist: Methodological ramifications of 'light' paternalism. In: Caplin, A., S. A. (Ed.), Handbook of Economic Methodologies. Oxford University Press.

Loewenstein, G., O'Donoghue, T., 2004. Animal spirits: Affective and deliberative processes in economic behavior. Available at SSRN 539843.

Loewenstein, G., Ubel, P. A., 2008. Hedonic adaptation and the role of decision and experience utility in public policy. Journal of Public Economics 92 (8), 1795–1810.

Loibl, C., Haisley, E., Jones, L., Loewenstein, G., 2012. Testing strategies to increase saving and retention in ida programs. Family and Consumers Science.

Long, R. T., 1998. Toward a libertarian theory of class. Social Philosophy and Policy 15 (02), 303–349.

Lotz, S., Baumert, A., Schlösser, T., Gresser, F., Fetchenhauer, D., 2011. Individual Differences in Third-Party Interventions: How Justice Sensitivity Shapes Altruistic Punishment. Negotiation and Conflict Management Research 4 (4), 297–313.

Luce, R. D., 2010. Interpersonal comparisons of utility for 2 of 3 types of people. Theory and decision 68 (1-2), 5–24.

Luo, H., Lu, S., Bharghavan, V., Cheng, J., Zhong, G., 2004. A packet scheduling approach to qos support in multihop wireless networks. Mobile Networks and Applications 9 (3), 193–206.

Mann, R. A., 1972. The behavior-therapeutic use of contingency contracting to control an adult behavior problem: Weight control. Journal of Applied Behavior Analysis.

Manski, C., 2000. Economic Analysis of Social Interactions. Journal of Economic Perspectives 14 (3), 115–136.

Marlowe, F., Berbesque, J., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J., Ensminger, J., Gurven, M., Gwako, E., Henrich, J., et al., 2008. More "Altruistic" Punishment in Larger Societies. Proceedings of the Royal Society B: Biological Sciences 275 (1634), 587–592.

Mas, A., Moretti, E., 2009. Peers at work. American Economic Review 99 (1), 112–145.

Mas-Colell, A., Whinston, M. D., Green, J. R., et al., 1995. Microeconomic theory. Vol. 1. New York: Oxford university press.

Mathew, S., Boyd, R., 2011. Punishment Sustains Large-scale Cooperation in Prestate Warfare. Proceedings of the National Academy of Sciences 108 (28), 11375–11380.

McGee, R. W., 2012. The Ethics of Tax Evasion: Perspectives in Theory and Practice. Springer, New York.

Mill, J. S., [1859]/2010. On liberty and other essays. Digireads. com Publishing.

Mills, G., Gale, W. G., Patterson, R., Engelhardt, G. V., Eriksen, M. D., Apostolov, E., 2008. Effects of individual development accounts on asset purchases and saving behavior: Evidence from a controlled experiment. Journal of Public Economics 92 (5), 1509–1530.

Mirrlees, J. A., 1971. An exploration in the theory of optimum income taxation. The review of economic studies 38 (2), 175–208.

Molm, L. D., 1994. Is Punishment Effective? Coercive Strategies in Social Exchange. Social Psychology Quarterly 57 (2), 75–94.

Morduch, J., 1999. The microfinance promise. Journal of economic literature 37 (4), 1569–1614.

Morgan, J., 2000. Financing Public Goods by means of Lotteries. The Review of Economic Studies 67 (4), 761–784.

Morgan, J., Sefton, M., 2000. Funding Public Goods with Lotteries: Experimental Evidence. The Review of Economic Studies 67 (4), 785–810.

Muller, D. C., 1989. Public choice II. New York: Cambridge University Press.

Musgrave, R. A., 1959. The Economic of Public Finance. MaGraw Hill New York.

Myerson, R. B., 1981. Utilitarianism, egalitarianism, and the timing effect in social choice problems. Econometrica: Journal of the Econometric Society, 883–897.

Nagel, T., 1970. The possibility of altruism. Oxford: Clarendon Press.

Nash, J. F., 1950. The bargaining problem. Econometrica: Journal of the Econometric Society, 155–162.

Nelissen, R., 2008. The Price You Pay: Cost-dependent Reputation Effects of Altruistic Punishment. Evolution and Human Behavior 29 (4), 242–248.

Nelissen, R., Zeelenberg, M., 2009. Moral Emotions as Determinants of Third-party Punishment: Anger, Guilt, and the Functions of Altruistic Sanctions. Judgment and Decision Making 4 (7), 543–553.

Newcomb, T. M., Koenig, K. E., Flacks, R., Warwick, D. P., 1967. Persistence and change: Bennington College and its students after twenty-five years. Wiley New York.

Ng, Y., 2007. Bentham or nash? on the acceptable form of social welfare functions. Economic Record 57 (3), 238–250.

Ng, Y.-K., 2000. Efficiency, equality and public policy: With a case for higher public spending. Hampshire (U.K.): Macmillan Press.

Nikiforakis, N., 2008. Punishment and Counter-punishment in Public Good Games: Can We Really Govern Ourselves? Journal of Public Economics 92 (1), 91–112.

Nozick, R., 1974. Anarchy, state, and utopia. Vol. 5038. Basic books.

Nussbaum, M. C., 2001. Women and human development: The capabilities approach. Vol. 3. Cambridge University Press.

Nuttin, J., Greenwald, A. G., 1968. Reward and Punishment in Human Learning: Elements of a Behavior Theory. Academic Press, New York.

O'Donoghue, T., Rabin, M., 1999. Doing it now or later. American Economic Review, 103–124.

O'Donoghue, T., Rabin, M., 2001. Choice and procrastination. The Quarterly Journal of Economics 116 (1), 121–160.

O'Donoghue, T., Rabin, M., 2003. Studying optimal paternalism, illustrated by a model of sin taxes. American Economic Review, 186–191.

Okimoto, T., Wenzel, M., 2011. Third-party Punishment and Symbolic Intragroup Status. Journal of Experimental Social Psychology 47 (4), 709–718.

Onderstal, S., Schram, A., Soetevent, A., 2011. Bidding to Give in the Field: Door-to-Door Fundraisers had it Right from the Start. Working paper, Tinbergen Institute.

Orzen, H., 2008. Fundraising through Competition: Evidence from the Lab. Working paper, CeDEx.

Ostrom, E., Walker, J., Gardner, R., 1992. Covenants with and without a sword: Self-governance is possible. American Political Science Review 86 (02), 404–417.

Parisi, F., 2000. Spontaneous Emergence of Law: Customary Law. In: Encyclopedia of Law and Economics. Edward Elgar Publishing, Camberley (UK).

Parisi, F., Rowley, C. K., 2005. The origins of law and economics: essays by the founding fathers. Northampton (USA): Edward Elgar Publishing.

Perkins, H., Linkenbach, J. W., Lewis, M. A., Neighbors, C., 2010. Effectiveness of social norms media marketing in reducing drinking and driving: A statewide campaign. Addictive behaviors 35 (10), 866–874.

Piazza, J., Bering, J., 2008. Concerns about Reputation via Gossip Promote Generous Allocations in an Economic Game. Evolution and Human Behavior 29 (3), 172–178.

Posner, E. A., 2000. Law and social norms: The case of tax compliance. Virginia Law Review, 1781–1819.

Posner, R. A., 1985. Wealth maximization revisited. Notre Dame Journal of Law, Ethics and Public Policy 2, 85.

Raihani, N., Grutter, A., Bshary, R., 2010. Punishers Benefit from Third-party Punishment in Fish. Science 327 (5962), 171–171.

Rawls, J., 1958. Justice as fairness. The philosophical review 67 (2), 164–194.

Rawls, J., 1971. A theory of justice. Cambridge: Harvard University Press.

Rawls, J., 1974a. The independence of moral theory. In: Proceedings and Addresses of the American Philosophical Association. Vol. 48. pp. 5–22.

Rawls, J., 1974b. Some reasons for the maximin criterion. The American Economic Review 64 (2), 141–146.

Rawls, J., 1985. Justice as fairness: political not metaphysical. Philosophy & Public Affairs 14 (3), 223–251.

Rawls, J., 2001. Justice as fairness: A restatement. Cambridge: Harvard University Press.

Rebonato, R., 2012. Taking Liberties: A Critical Examination of Libertarian Paternalism. Palgrave Macmillan.

Rege, M., Telle, K., 2004. The impact of social approval and framing on cooperation in public good situations. Journal of public Economics 88 (7), 1625–1644.

Robbins, L., [1932 Or. Ed.]-2007. An essay on the nature and significance of economic science. Ludwig von Mises Institute.

Roemer, J. E., 1998. Theories of distributive justice. Harvard University Press.

Samuelson, P. A., 1938. A note on the pure theory of consumer's behaviour. Economica, 61–71.

Samuelson, P. A., 1947. Foundations of economic analysis. Cambridge: Harvard University Press.

Scharfstein, D., Stein, J., 1990. Herd Behavior and Investment. The American Economic Review, 465–479.

Schram, A., Onderstal, S., 2009. Bidding to Give: an Experimental Comparison of Auctions for Charity. International Economic Review 50 (2), 431–457.

Schreiner, M., Sherraden, M. W., 2007. Can the poor save?: saving & asset building in individual development accounts. Transaction Publishers.

Sen, A., 1973. On economic inequality. Oxford: Clarendon Press.

Sen, A., 1977. Social choice theory: A re-examination. Econometrica: journal of the Econometric Society, 53–89.

Sen, A., 1980. Equality of what? The Tanner lectures on human values 1, 353–369.

Sen, A., 1985. Commodities and capabilities. Professor Dr. P. Hennipman lectures in economics: theory, institutions, policy (7.

Sen, A., 1999. Development as freedom. Oxford University Press.

Sen, A., et al., 1993. Capability and well-being. The quality of life 1 (9), 30–54.

Sen, A. K., 1970. Collective choice and social welfare. Amsterdam: North-Holland Publishing Co.

Sen, A. K., 1997. Choice, welfare, and measurement. Cambridge: Harvard University Press.

Sen, A. K., 2009. The idea of justice. Cambridge: Harvard University Press.

Seymour, B., Singer, T., Dolan, R., 2007. The Neurobiology of Punishment. Nature Reviews Neuroscience 8 (4), 300–311.

Shinada, M., Yamagishi, T., Ohmura, Y., 2004. False Friends are Worse than Bitter Enemies:"Altruistic" Punishment of In-group Members. Evolution and Human Behavior 25 (6), 379–393.

Sims, H. P., 1980. Further Thoughts on Punishment in Organizations. Academy of Management Review 5 (1), 133–138.

Slemrod, J., 2007. Cheating Ourselves: The Economics of Tax Evasion. The Journal of Economic Perspectives 21 (1), 25–48.

Slemrod, J., Yitzhaki, S., 2002. Tax Avoidance, Evasion, and Administration. Handbook of public economics 3, 1423–1470.

Spector, H., 2004. Fairness and welfare from a comparative law perspective. Chi.-Kent L. Rev. 79, 521.

Stigler, G. J., 1971. The Theory of Economic Regulation. Bell Journal of Economics 2 (1), 3–21.

Sugden, R., 2004. The opportunity criterion: consumer sovereignty without the assumption of coherent preferences. American Economic Review, 1014–1033.

Sugden, R., 2009. On nudging: A review of nudge: Improving decisions about health, wealth and happiness by richard h. thaler and cass r. sunstein. International Journal of the Economics of Business 16 (3), 365–373.

Sunstein, C. R., 2013. The storrs lectures: behavioral economics and paternalism. Yale LJ 122, 1826–2082.

Sunstein, C. R., Schkade, D., Ellman, L. M., Sawick, A., 2006. Are judges political?: an empirical analysis of the federal judiciary. Brookings Institution Press.

Sunstein, C. R., Thaler, R. H., 2003. Libertarian paternalism is not an oxymoron. The University of Chicago Law Review, 1159–1202.

Swope, K., Cadigan, J., Schmitt, P., Shupp, R., 2008. Social position and distributive justice: Experimental evidence. Southern Economic Journal, 811–818.

Terrab, M., Odoni, A. R., 1993. Strategic flow management for air traffic control. Operations Research 41 (1), 138–152.

Thaler, R. H., Benartzi, S., 2004. Save more tomorrow$^{\text{TM}}$: Using behavioral economics to increase employee saving. Journal of political Economy 112 (S1), S164–S187.

Thaler, R. H., Sunstein, C. R., 2003. Libertarian paternalism. American Economic Review, 175–179.

Thaler, R. H., Sunstein, C. R., 2008. Nudge: Improving decisions about health, wealth, and happiness. Yale University Press.

Tobin, J., 1970. On limiting the domain of inequality. Journal of Law and Economics 13 (2), 263–77.

Topa, G., 2001. Social Interactions, Local Spillovers and Unemployment. The Review of Economic Studies 68 (2), 261–295.

Torgler, B., 2003. Beyond Punishment: A Tax Compliance Experiment with Taxpayers in Costa Rica. Revista de Análisis Económico 18 (1).

Tullock, G., Brady, G. L., Seldon, A., 2002. Government Failure: A Primer in Public Choice. Cato Institute.

Turner, J. C., 1991. Social influence. Thomson Brooks/Cole Publishing Co.

Tversky, A., Kahneman, D., 1981. The framing of decisions and the psychology of choice. Science 211 (4481), 453–458.

Tversky, A., Kahneman, D., 1986. Rational choice and the framing of decisions. Journal of business, S251–S278.

Tversky, A., Kahneman, D., 1992. Advances in Prospect Theory: Cumulative Representation of Uncertainty. Journal of Risk and uncertainty 5 (4), 297–323.

Ubel, P. A., Loewenstein, G., Schwarz, N., Smith, D., 2005. Misimagining the unimaginable: the disability paradox and health care decision making. Health Psychology 24 (4S), S57.

Van Praag, B. M., 1993. The relativity of the welfare concept. In: Nussbaum, M., Sen, A. (Eds.), The quality of life. Oxford University Press, pp. 200–254.

Volpp, K. G., John, L. K., Troxel, A. B., Norton, L., Fassbender, J., Loewenstein, G., 2008. Financial incentive-based approaches for weight loss: A randomized trial. JAMA, the journal of the American Medical Association 300 (22), 2631–2637.

Volpp, K. G., Troxel, A. B., Pauly, M. V., Glick, H. A., Puig, A., Asch, D. A., Galvin, R., Zhu, J., Wan, F., DeGuzman, J., et al., 2009. A randomized, controlled trial of financial incentives for smoking cessation. New England Journal of Medicine 360 (7), 699–709.

Von Neumann, J., Morgenstern, O., 1944. The Theory of Games and Economic Behavior. Princeton University Press, Princeton.

Walzer, M., 1983. Sphere of Justice: A Defense of Pluralism andEquality. Basic Books New York.

Wan, J., 2010. The Incentive to Declare Taxes and Tax Revenue: the Lottery Receipt Experiment in China. Review of Development Economics 14 (3), 611–624.

Webley, P., Adams, C., Elffers, H., 2006. Value Added Tax Compliance. In: MacCaffrey, E. J., Slemrod, J. B. (Eds.), Behavioral Public Finance. Russell Sage Foundation, New York, pp. 175–205.

Wright, J. D., Ginsburg, D. H., 2012. Behavioral law and economics: Its origins, fatal flaws, and implications for liberty. Nw. UL Rev. 106, 1033.

Yitzhaki, S., 1974. Income Tax Evasion: A Theoretical Analysis. Journal of Public Economics 3 (2), 201–202.

Young, H. P., 1995. Equity: in theory and practice. Princeton: Princeton University Press.

Zasu, Y., 2007. Sanctions by Social Norms and the Law: Substitutes or Complements? The Journal of Legal Studies 36 (2), 379–396.